Generalised Hunter–Saxton equations and optimal information transport

Klas MODIN[1,2,*]

[1] *Department of Mathematics*
University of Toronto
Toronto, ON M5S 2E4, Canada

[2] *Department of Mathematical Sciences*
Chalmers University of Technology
SE–412 96 Göteborg, Sweden

February 27, 2013

A right invariant H^1–type Riemannian metric on the group of diffeomorphisms of a compact manifold is studied. The significance of this metric is that it descends, by a Riemannian submersion, to the constant curvature Fisher metric on the space of smooth probability densities.

The right reduced geodesic equation is a higher dimensional generalisation of the μ–Hunter–Saxton equation, which describes liquid crystals under influence of an external magnetic field. A local existence and uniqueness result is obtained by proving that the geodesic spray is smooth with respect to H^s Banach topologies.

Based on the descending property of the metric, a polar factorisation result for diffeomorphisms is given. Analogous to the polar factorisation used in optimal mass transport, this factorisation solves a corresponding optimal information transport problem with respect to the Fisher metric. It can be seen as an infinite dimensional version of the classical QR factorisation of matrices.

Keywords: Euler–Arnold equations, Euler–Poincare equations, Cholesky factorisation, descending metrics, diffeomorphism groups, Fisher metric, geometric statistics, Hunter–Saxton equation, information geometry, optimal transport, polar factorisation, QR factorisation, Riemannian submersion.

MSC 2010: 58D05, 58D15, 35Q31, 53C21, 58B20, 94A17, 65F99.

*klas.modin@chalmers.se

Contents

1 Introduction

Let (M, g) be a closed Riemannian manifold of dimension n, and denote the induced volume form by vol. We assume that vol is normalised, i.e., that $\int_M \mathrm{vol} = 1$. The group of diffeomorphisms of M is denoted $\mathrm{Diff}(M)$. The subgroup of volume preserving diffeomorphisms is denoted $\mathrm{Diff}_{\mathrm{vol}}(M)$. Throughout the paper, the word “metric” always refers to “Riemannian metric”.

The study of geodesic equations on diffeomorphism groups was initiated by Arnold [2], who discovered that the Euler equations of an incompressible perfect fluid correspond to a geodesic equation on $\mathrm{Diff}_{\mathrm{vol}}(M)$ with respect to a right invariant L^2 metric. Since Arnold's discovery, it has been shown that many equations of mathematical physics can be put into this framework. (Such equations are often called *Euler–Poincaré equations* or *Euler–Arnold equations*. See the monographs [3, 24, 20, 15] and the survey paper [34].)

The subject of this paper concerns geodesic equations corresponding to a class of right invariant Riemannian metrics on $\mathrm{Diff}(M)$. The significance of these metrics is that they descend to the homogeneous space $\mathrm{Diff}_{\mathrm{vol}}(M)\backslash\mathrm{Diff}(M)$ of right co-sets, naturally identified with the space $\mathrm{Dens}(M)$ of smooth probability densities.

Riemannian metrics and geodesic equations on $\mathrm{Dens}(M)$ are of importance in optimal transport, probability theory, statistical mechanics, and quantum mechanics. The connection between geodesics on $\mathrm{Diff}(M)$ and $\mathrm{Dens}(M)$ was pointed out by Otto [29], who studied a non-invariant L^2 metric on $\mathrm{Diff}(M)$ which descends to $\mathrm{Dens}(M)$. (In this setting, $\mathrm{Dens}(M)$ is identified with the homogeneous space $\mathrm{Diff}(M)/\mathrm{Diff}_{\mathrm{vol}}(M)$ of *left* co-sets.) Remarkably, the corresponding metric on $\mathrm{Dens}(M)$ induces the L^2 Wasserstein distance, and is therefore called the *Wasserstein metric*. Otto's observation implies that the L^2 optimal mass transport problem, which belongs to the class of Monge–Kantorovich problems, can be interpreted as a geodesic problem on $\mathrm{Dens}(M)$ with respect to the Wasserstein metric (at least in the case of smooth densities). This follows from general facts about Riemannian submersions: (i) a minimal geodesic between two fibres must be horizontal, and (ii) a horizontal geodesic descends to a geodesic on the base.

Another important metric on the space of probability densities is the *Fisher metric* (also called the Fisher–Rao, or entropy differential metric). Classically, the Fisher metric occurs as a finite dimensional metric on smooth statistical models (called statistical manifolds), and has fundamental rôle in the field of information geometry [11, 31, 5, 1]. Friedrich [12] realised that statistical manifolds can be interpreted as finite dimensional submanifolds of $\mathrm{Dens}(M)$, and that the Fisher metric on such statistical manifolds is the restriction of one and the same canonical Fisher metric on $\mathrm{Dens}(M)$. Furthermore, Friedrich showed that this metric has constant positive curvature.

Khesin, Lenells, Misiołek, and Preston [19] introduced a right invariant degenerate $\dot{H}^1$ "metric" on $\mathrm{Diff}(M)$, which descends to the Fisher metric on $\mathrm{Dens}(M)$. ($\mathrm{Dens}(M)$ is here identified with the *right* co-sets $\mathrm{Diff}_{\mathrm{vol}}(M)\backslash\mathrm{Diff}(M)$.) By taking Otto's point of view, the authors then regard the geodesic problem, with respect to the Fisher metric, as an optimal information transport problem, with respect to a degenerate cost function on $\mathrm{Diff}(M)$ induced by the $\dot{H}^1$ "metric". Since the cost function is degenerate, the solutions are not unique.

Ideally, one would like to have a right invariant metric on $\mathrm{Diff}(M)$ which descends to the Fisher metric on $\mathrm{Dens}(M)$. It is remarked in [19] that examples of non-degenerate right invariant metrics on $\mathrm{Diff}(M)$ descending to $\mathrm{Dens}(M)$ are lacking. The main motivation for the work at hand is to construct such metrics, and thus complete the analogy between optimal mass transport and optimal information transport.

Indeed, in this paper we introduce a 3–parameter family of right invariant metrics on $\mathrm{Diff}(M)$ descending to the Fisher metric, and we prove local existence and uniqueness of the geodesic equations. We also attain existence and uniqueness for the corresponding optimal information transport problem, which, in turn, implies a polar factorisation result for diffeomorphisms. This factorisation is analogous to the polar factorisation of vector valued maps on $\mathbb{R}^n$, obtained by Brenier [4], and later generalised to manifolds by McCann [25]. Our factorisation result can be seen as an infinite dimensional version of the classical QR factorisation of matrices.

The right reduced geodesic equations for the family of metrics, which we are now going to present, can be interpreted as higher dimensional generalisations of the μ–Hunter–Saxton (μHS) equation, introduced by Khesin, Lenells, and Misiołek [18] (also called μ–Camassa–Holm in [23]). μHS is a simple model for liquid crystals under influence of an external magnetic field.

Let $\mathfrak{X}(M)$ denote the smooth vector fields and $\Omega^k(M)$ the smooth k–forms on M. Further, let

$$\mathcal{F}(M) = \left\{ F \in C^\infty(M); \int_M F\,\mathrm{vol} = 0 \right\}.$$

Recall the differential $\mathrm{d} : \Omega^k(M) \to \Omega^{k+1}(M)$, and the co-differential $\delta : \Omega^k(M) \to \Omega^{k-1}(M)$. The Laplace–de Rham operator $\Delta = -\mathrm{d}\circ\delta - \delta\circ\mathrm{d}$ restricted to $\mathrm{d}\Omega^{k-1}(M)$ or $\delta\Omega^{k+1}(M)$ is an isomorphism [32]. In particular, it is an isomorphism on $\mathcal{F}(M) = \delta\Omega^1(M)$. Let $\flat : \mathfrak{X}(M) \to \Omega^1(M)$ denote the flat map, also called the *musical isomorphism*. Its inverse, the sharp map, is denoted $\sharp$. For $u \in \mathfrak{X}(M)$, we write $u^\flat$ instead of $\flat(u)$ and correspondingly for $\sharp$.

Consider the pseudo-differential operator $\mathcal{A} : \mathfrak{X}(M) \to \Omega^1(M)$ defined by

$$\mathcal{A}u := \Big(\mathrm{id} + \mathrm{d} \circ \Delta^{-1} \circ \delta + \gamma\, \delta \circ \Delta^{-1} \circ \mathrm{d} + \alpha\, \delta \circ \mathrm{d} + \beta\, \mathrm{d} \circ \delta\Big)(u^\flat) \tag{1}$$

where $\alpha, \beta > 0$ and $\gamma \in [0,1]$ are parameters. We are interested in the integro-differential equation given by

$$\dot{m} + \mathcal{L}_u m + m\,\mathrm{div}(u) = 0, \quad m = \mathcal{A}u, \tag{2a}$$

where $\mathcal{L}_u$ denotes the Lie derivative along u and $\dot{m} = \frac{\partial m}{\partial t}$. By a solution we mean a curve $t \mapsto u(t) \in \mathfrak{X}(M)$ such that u fulfils equation (2a). The equation also admits the form

$$\Big(\frac{\partial}{\partial t} + \mathcal{L}_u\Big)(m \otimes \mathrm{vol}) = 0, \tag{2b}$$

which follows since

$$\mathcal{L}_u(m \otimes \mathrm{vol}) = (\mathcal{L}_u m) \otimes \mathrm{vol} + m \otimes \mathrm{div}(u)\mathrm{vol} = \big(\mathcal{L}_u m + m\,\mathrm{div}(u)\big) \otimes \mathrm{vol}.$$

The paper is organised as follows. In § 2 we show that equation (2) is a right reduced equation for geodesics on Diff(M), i.e., an Euler–Arnold equation. Local existence and uniqueness of the Cauchy problem is given in § 3. In § 4 we discuss characterisation and construction of right invariant and descending metrics, and we show that the family of metrics constructed in this paper descend to the Fisher metric. In § 5 we first consider an abstract geometric framework for right invariant optimal transport problems and polar factorisation. Then, in § 5.1, we focus on the case of optimal information transport using the new metric, and we derive a polar factorisation result for H^s diffeomorphisms. Finally, we show in § 5.2 that the classical QR factorisation of matrices can be viewed as a polar factorisation corresponding to optimal transport of inner products on $\mathbb{R}^n$. The relation to the Cholesky factorisation of symmetric matrices is also pointed out.

Before this, we continue below with a derivation of yet another form of equation (2), based on the Hodge decomposition. This form reveals some structural properties and relation to other equations.

1.1 Hodge components

From the Helmholtz decomposition it follows that $\mathfrak{X}(M) = \mathfrak{X}_{\mathrm{vol}}(M) \oplus \mathrm{grad}(\mathcal{F}(M))$, where $\mathfrak{X}_{\mathrm{vol}}(M)$ denotes the divergence free vector fields. Hence, every $u \in \mathfrak{X}(M)$ can be decomposed uniquely as $u = \xi + \mathrm{grad}(f)$, with $\xi \in \mathfrak{X}_{\mathrm{vol}}(M)$ and $f \in \mathcal{F}(M)$. Notice that f is unique, since it is required to be normalised. This is an orthogonal decomposition with respect to the L^2 inner product on $\mathfrak{X}(M)$, which is given by

$$\langle u, v\rangle_{L^2} = \int_M \mathsf{g}(u,v)\mathrm{vol}.$$

For k–forms, the Hodge decomposition is given by

$$\Omega^k(M) = \mathcal{H}^k(M) \oplus \delta\Omega^{k+1}(M) \oplus \mathrm{d}\Omega^{k-1}(M),$$

where $\mathcal{H}^k(M) = \{a \in \Omega^k(M); \Delta a = 0\}$ is the space of harmonic k–forms. This decomposition is orthogonal with respect to the L^2 inner product on $\Omega^k(M)$, which is given by

$$\langle a, b\rangle_{L^2} = \int_M a \wedge \star b,$$

where $\star : \Omega^k(M) \to \Omega^{n-k}(M)$ is the Hodge star map. Notice that $\langle u, v\rangle_{L^2} = \langle u^\flat, v^\flat\rangle_{L^2}$.

Let $\mathsf{D}^k(M) = \mathcal{H}^k(M) \oplus \delta\Omega^{k+1}(M)$. Then $\mathsf{D}^k(M) = \{a \in \Omega^k(M); \delta a = 0\}$ is the space of co-closed k–forms. The relation between the Hodge decomposition and the Helmholtz decomposition is:

$$\mathfrak{X}_{\mathrm{vol}}(M)^\flat = \mathsf{D}^1(M), \quad \mathrm{grad}(\mathcal{F}(M))^\flat = \mathrm{d}\Omega^0(M).$$

In other words, the musical isomorphism $\flat : \mathfrak{X}(M) \to \Omega^1(M)$ is diagonal with respect to the Helmholtz and Hodge decompositions. The same holds for the pseudo differential operator $\mathcal{A} : \mathfrak{X}(M) \to \Omega^1(M)$. That is,

$$\mathcal{A}\mathfrak{X}_{\mathrm{vol}}(M) = \mathsf{D}^1(M), \quad \mathcal{A}\,\mathrm{grad}(\mathcal{F}(M)) = \mathrm{d}\Omega^0(M).$$

From the Hodge decomposition we also obtain a finer decomposition

$$\mathfrak{X}_{\mathrm{vol}}(M) = \mathfrak{X}_{\mathcal{H}}(M) \oplus \mathfrak{X}_{\mathrm{vol,ex}}(M),$$

where $\mathfrak{X}_{\mathrm{vol,ex}}(M) = \delta\Omega^2(M)^\sharp$ are the *exact* volume preserving vector fields, and $\mathfrak{X}_{\mathcal{H}}(M) = \mathcal{H}^1(M)^\sharp$ are the *harmonic* vector fields. $\mathcal{A}$ is also diagonal with respect to this finer decomposition. Indeed, the L^2 orthogonal projection operator $R : \Omega^1(M) \to \mathcal{H}^1(M)$ onto the harmonic part is given by

$$R = \mathrm{id} + \mathrm{d} \circ \Delta^{-1} \circ \delta + \delta \circ \Delta^{-1} \circ \mathrm{d},$$

and the L^2 orthogonal projection operator $P : \Omega^1(M) \to \mathsf{D}^1(M)$ onto the co-closed part is given by

$$P = \mathrm{id} + \mathrm{d} \circ \Delta^{-1} \circ \delta.$$

From the definition of $\mathcal{A}$ it follows that $\mathcal{A} = (\gamma R + (1-\gamma)P + \alpha\,\delta \circ \mathrm{d} + \beta\,\mathrm{d} \circ \delta) \circ \flat$. Now, if $h \in \mathfrak{X}_{\mathcal{H}}(M)$ then

$$\mathcal{A}h = \underbrace{\big(\gamma R + (1-\gamma)P\big)h^\flat}_{h^\flat} + \alpha\delta\,\underbrace{\mathrm{d}h^\flat}_{0} + \beta\mathrm{d}\,\underbrace{\delta h^\flat}_{0} = h^\flat \in \mathcal{H}^1(M),$$

if $\xi \in \mathfrak{X}_{\mathrm{vol,ex}}(M)$ then

$$\mathcal{A}\xi = \underbrace{\gamma R\xi^\flat}_{0} + (1-\gamma)\,\underbrace{P\xi^\flat}_{\xi^\flat} + \alpha\delta\mathrm{d}\xi^\flat + \beta\mathrm{d}\,\underbrace{\delta\xi^\flat}_{0} = (1-\gamma)\xi^\flat + \alpha\delta\mathrm{d}\xi^\flat \in \delta\Omega^2(M),$$

and

$$\mathcal{A}\,\mathrm{grad}(f) = \underbrace{(1-\gamma)P\mathrm{d}f + \gamma R\mathrm{d}f + \alpha\,\delta\mathrm{d}\mathrm{d}f}_{0} - \beta\,\mathrm{d}\Delta f = -\beta\,\mathrm{d}\Delta f \in \mathrm{d}\Omega^0(M).$$

Thus, if we represent $u = h + \xi + \operatorname{grad}(f)$ by its unique "Helmholtz–Hodge components" $(h, \xi, f) \in \mathfrak{X}_{\mathcal{H}}(M) \times \mathfrak{X}_{\mathrm{vol,ex}}(M) \times \mathcal{F}(M)$, we have that

$$\mathcal{A}(h, \xi, f) = (h^\flat, \big((1-\gamma)\mathrm{id} - \alpha\Delta\big)\xi^\flat, -\beta\Delta f) \in \mathcal{H}^1(M) \times \delta\Omega^2(M) \times \mathcal{F}(M).$$

Since both $\big((1-\gamma)\mathrm{id} - \alpha\Delta\big) \circ \flat : \mathfrak{X}_{\mathrm{vol,ex}}(M) \to \delta\Omega^2(M)$ and $\Delta : \mathcal{F}(M) \to \mathcal{F}(M)$ are invertible operators, it follows that $\mathcal{A}$ is also invertible (see § 3 for details).

Our aim is now to write equation (2a) in terms of the Hodge components

$$\sigma := (\gamma R + (1-\gamma)P - \alpha\, \delta \circ \mathrm{d})(u^\flat) \in \mathsf{D}^1(M)$$

and

$$\rho := \Delta f = \operatorname{div}(u) \in \mathcal{F}(M),$$

corresponding to the partial Hodge decomposition $\Omega^1(M) = \mathsf{D}^1(M) \oplus \mathrm{d}\mathcal{F}(M)$. In these variables $m = \sigma - \beta \mathrm{d}\rho$, so equation (2a) becomes

$$\dot\sigma - \beta\,\mathrm{d}\dot\rho + \mathcal{L}_u\sigma - \beta\,\mathrm{d}\mathcal{L}_u\rho + \rho\sigma - \beta\rho\,\mathrm{d}\rho = 0$$
$$\Updownarrow$$
$$\dot\sigma + \mathcal{L}_u\sigma + \rho\sigma - \beta\,\mathrm{d}\Big(\dot\rho + \mathcal{L}_u\rho + \frac{\rho^2}{2}\Big) = 0$$

Notice, in general, that $\mathcal{L}_u\sigma + \rho\sigma \notin \mathsf{D}^1(M)$ and that $\mathcal{L}_u\rho + \frac{\rho^2}{2} \notin \mathcal{F}(M)$. Thus, in order to find the Hodge components, we have to introduce a Lagrangian multiplier $p \in C^\infty(M)$. We can always find a p such that $\mathcal{L}_u\xi^\flat + \rho\xi^\flat + \mathrm{d}p \in \mathsf{D}^1(M)$, and such a p is uniquely determined up to a constant. Further, we can always determine the constant part of p in such a way that $\mathcal{L}_u\rho + \frac{\rho^2}{2} + \frac{p}{\beta} \in \mathcal{F}(M)$. Continuing from above

$$\Updownarrow$$
$$\underbrace{\dot\sigma + \mathcal{L}_u\sigma + \rho\sigma + \mathrm{d}p}_{\in \mathsf{D}^1(M)} - \beta\,\mathrm{d}\Big(\underbrace{\dot\rho + \mathcal{L}_u\rho + \frac{\rho^2}{2} + \frac{p}{\beta}}_{\in\mathcal{F}(M)}\Big) = 0.$$

We now obtain equation (2a) in terms of the Hodge components as

$$\begin{aligned}
\dot\sigma + \mathcal{L}_u\sigma + \rho\sigma &= -\mathrm{d}p, \qquad \sigma = \big(\gamma R + (1-\gamma)\mathrm{id} - \alpha\Delta\big)(Pu^\flat) \\
\dot\rho + \mathcal{L}_u\rho + \frac{\rho^2}{2} &= -\frac{p}{\beta}, \qquad \rho = \operatorname{div}(u) \\
\delta\sigma &= 0 \\
\int_M \rho\,\mathrm{vol} &= 0,
\end{aligned} \tag{2c}$$

where the "pressure" $p \in C^\infty(M)$ is a Lagrangian multiplier, determined uniquely by the two constraint equations.

Notice from equation (2c) that if $\sigma(t_0) = 0$ at some time t_0, then $\dot\sigma(t_0) = 0$. As a consequence, $\mathrm{grad}(\mathcal{F}(M))$ is an *invariant subspace* of equation (2), so if $u(t_0) \in \mathrm{grad}(\mathcal{F}(M))$ then $u(t) \in \mathrm{grad}(\mathcal{F}(M))$ for all t. From a geometric point of view, the reason for this is that the corresponding right invariant metric on $\mathrm{Diff}(M)$ descends to the homogenous space $\mathrm{Diff}_{\mathrm{vol}}(M)\backslash\mathrm{Diff}(M) \simeq \mathrm{Dens}(M)$, as is described in § 4. In contrast, $\rho(t_0) = 0$ does *not* imply that $\rho(t) = 0$, so $\mathfrak{X}_{\mathrm{vol}}(M)$ is not an invariant subspace. However, if $\rho(t_0) = 0$ then it follows from equation (2c) that $\dot\rho(t_0)$ is arbitrarily small for large enough β. In the case $\gamma = 0$, this observation suggests that solutions to equation (2) may converge to solutions of the Euler–α fluid equation as $\beta \to \infty$, which is to be investigated in future work. We do not expect good behaviour of solutions as $\beta \to 0$, since $\mathcal{A}$ is not invertible for $\beta = 0$.

Equation (2) is a higher dimensional generalisation of the μHS equation, studied by Khesin, Lenells, and Misiołek [18]. Indeed, if $M = S^1$ then $\mathfrak{X}_{\mathrm{vol}}(S^1) = \mathfrak{X}_{\mathcal{H}}(S^1) \simeq \mathbb{R}$ consists of the constant vector fields on S^1. Equation (2c) then becomes

$$\begin{aligned} \dot\xi + 2\xi u_x &= -p_x \\ \dot u_x + u u_{xx} + \frac{1}{2}(u_x)^2 &= -\frac{p}{\beta}. \end{aligned}$$

From the first equation it follows that

$$0 = \int_{S^1} (\dot\xi + 2\xi u_x + p_x)\,\mathrm{d}x = \dot\xi \mu(S^1) \qquad \text{where} \qquad \mu(S^1) = \int_{S^1} \mathrm{d}x$$

which implies that $\dot\xi = 0$. If the second equation is differentiated with respect to x we get

$$\dot u_{xx} + 2u_x u_{xx} + u u_{xxx} = \frac{2\xi u_x}{\beta}.$$

Since $\int_{S^1} u\,\mathrm{d}x = \int_{S^1} \xi\,\mathrm{d}x$ it follows that ξ is the mean of u over S^1, i.e.,

$$\xi = \mu(u) := \frac{1}{\mu(S^1)} \int_{S^1} u\,\mathrm{d}x.$$

Thus, we finally arrive at

$$\dot u_{xx} + 2u_x u_{xx} + u u_{xxx} = \frac{2\mu(u) u_x}{\beta}$$

which is the μHS equation.

A different generalisation of μHS, from $M = S^1$ to $M = \mathbb{T}^n$ (the n–dimensional flat torus), is given by Kohlmann [21]. The equation suggested there is also an Euler–Arnold equation, hence a geodesic equation on $\mathrm{Diff}(\mathbb{T}^n)$. The corresponding right invariant metric does not descend to density space.

2 Euler–Arnold structure

The geodesic equation for a right (or left) invariant metric on a Lie group G can be reduced to an equation on the Lie algebra $\mathfrak{g}$, called an Euler–Poincaré or Euler–Arnold equation. The abstract form of this equation, first written down by Poincaré [30], is

$$\mathcal{A}\dot{u} + \mathrm{ad}_u^*(\mathcal{A}u) = 0, \tag{3}$$

where $\mathcal{A} : \mathfrak{g} \to \mathfrak{g}^*$ is the *inertia operator* induced by the inner product on $\mathfrak{g}$ corresponding to the right invariant metric, and $\mathrm{ad}_u^* : \mathfrak{g}^* \to \mathfrak{g}^*$ is the infinitesimal action of u on $\mathfrak{g}^*$, i.e., the dual operator of $\mathrm{ad}_u : \mathfrak{g} \to \mathfrak{g}$.

In our case, $G = \mathrm{Diff}(M)$, $\mathfrak{g} = \mathfrak{X}(M)$ and $\mathrm{ad}_u = -\mathcal{L}_u$, i.e., minus the Lie derivative (acting on vector fields). We identify the dual of $\mathfrak{X}(M)$ with $\Omega^1(M)$ via the pairing

$$\langle m, u \rangle = \int_M \mathrm{i}_u m \,\mathrm{vol} = \langle m, v^\flat \rangle_{L^2}.$$

Next, we introduce an inner product on $\mathfrak{X}(M)$, whose inertia operator is given by (1). Indeed, a non-degenerate inner product on $\mathfrak{X}(M)$ is given by

$$\langle u, v \rangle_{\alpha\beta\gamma} = \langle \bar{P}_\gamma u^\flat, \bar{P}_\gamma v^\flat \rangle_{L^2} + \alpha \langle \mathrm{d}u^\flat, \mathrm{d}v^\flat \rangle_{L^2} + \beta \langle \delta u^\flat, \delta v^\flat \rangle_{L^2}, \tag{4}$$

where $\bar{P}_\gamma = \gamma R + (1-\gamma) P$ is introduced to simplify the notation. Notice that (4) is different from the Sobolev *a-b-c* inner product, recently considered in [19], since only the divergence free components occur in the first term. By using that $\langle a, \delta b \rangle_{L^2} = \langle \mathrm{d}a, b \rangle_{L^2}$ and $\langle \bar{P}_\gamma u^\flat, \bar{P}_\gamma v^\flat \rangle_{L^2} = \langle \bar{P}_\gamma u^\flat, v^\flat \rangle_{L^2}$ we get

$$\langle u, v \rangle_{\alpha\beta\gamma} = \langle \bar{P}_\gamma u^\flat + \alpha\, \delta \mathrm{d}u^\flat + \beta\, \mathrm{d}\delta u^\flat, v^\flat \rangle_{L^2} = \langle \mathcal{A}u, v \rangle,$$

so $\mathcal{A}$ in (1) is the inertia tensor corresponding to the inner product (4).

Using this inner product, we define a right invariant metric $\langle\!\langle \cdot, \cdot \rangle\!\rangle_{\alpha\beta\gamma}$ on $\mathrm{Diff}(M)$ by right translation of vectors to $\mathfrak{X}(M) = T_{\mathrm{id}}\mathrm{Diff}(M)$. Explicitly,

$$\langle\!\langle U, V \rangle\!\rangle_{\alpha\beta\gamma} = \langle U \circ \varphi^{-1}, V \circ \varphi^{-1} \rangle_{\alpha\beta\gamma} \,, \tag{5}$$

for $U, V \in T_\varphi \mathrm{Diff}(M)$.

In the special case $M = S^1$ we have that $\mathrm{Diff}_{\mathrm{vol}}(S^1) = \mathrm{Rot}(S^1)$, i.e., the one dimensional manifold of rigid rotations. Thus, $\mathfrak{X}_{\mathrm{vol}}(S^1) \simeq \mathbb{R}$ consists of the constant vector fields on S^1. The inner product (4) on $\mathfrak{X}(S^1)$ then becomes

$$\langle u, v \rangle_{\mu\dot{H}^1} = \int_{S^1} u \,\mathrm{d}s \int_{S^1} v \,\mathrm{d}s + \int_{S^1} u_x v_x \,\mathrm{d}s$$

which is exactly the inner product defining the μHS metric. Therefore, the α-β-γ metric (5) on $\mathrm{Diff}(M)$ is a generalisation of the μHS metric on $\mathrm{Diff}(S^1)$ to arbitrary compact manifolds.

The dual operator of $-\mathfrak{L}_u : \mathfrak{X}(M) \to \mathfrak{X}(M)$ is computed as

$$\begin{aligned}
\langle m, -\mathfrak{L}_u v\rangle &= -\int_M m \wedge \star(\mathfrak{L}_u v)^\flat = -\int_M m \wedge \mathrm{i}_{\mathfrak{L}_u v}\mathrm{vol} \\
&= -\int_M m \wedge \big(\mathfrak{L}_u \mathrm{i}_v \mathrm{vol} - \mathrm{i}_v \mathfrak{L}_u \mathrm{vol}\big) \\
&= \underbrace{-\int_M \mathfrak{L}_u\big(m \wedge \mathrm{i}_v\mathrm{vol}\big)}_{0} + \int_M \mathfrak{L}_u m \wedge \big(\mathrm{i}_v \mathrm{vol} + \mathrm{div}(u)\mathrm{i}_v\mathrm{vol}\big) \\
&= \int_M \big(\mathfrak{L}_u m + \mathrm{div}(u) m\big) \wedge \mathrm{i}_v\mathrm{vol} = \langle \mathfrak{L}_u m + \mathrm{div}(u)m, v\rangle.
\end{aligned}$$

As a result, $\mathrm{ad}_u^*(m) = \mathfrak{L}_u m + \mathrm{div}(u)m$. From (3) we then obtain the following.

Proposition 2.1. *Equation (2) is the Euler–Arnold equation for the geodesic flow on* $\mathrm{Diff}(M)$ *with respect to the right invariant α-β-γ metric (5).*

3 Local existence and uniqueness

In this section we show that equation (2) is well posed as a Cauchy problem. The approach is that of Ebin and Marsden [7], which is to prove that the geodesic spray corresponding to the α-β-γ metric (5) is smooth with respect to Sobolev H^s topologies.

Let N be a smooth finite dimensional manifold. If $s > n/2$ then the set $H^s(M,N)$ of maps from M to N of Sobolev differentiability H^s is a Banach manifold (in fact, $H^s(M,N)$ is a Hilbert manifold, but that is not relavant for our analysis). Let $\pi_N : TN \to N$ be the canonical projection. The tangent space at $f \in H^s(M,N)$ is given by

$$T_f H^s(M,N) = \{v \in H^s(M,TN); \pi_N \circ v = f\}.$$

Thus, $TH^s(M,N) = H^s(M,TN)$. By iterating this we obtain, for higher order tangent spaces, that $T^k H^s(M,N) = H(M,T^kN)$.

If $s > n/2 + 1$, which we assume throughout the remainder, then $\mathrm{Diff}^s(M)$, i.e., the set of bijective maps in $H^s(M,M)$ whose inverses also belong to $H^s(M,M)$, is an open subset of $H^s(M,M)$, and therefore also a Banach manifold. Since $\mathrm{Diff}^s(M)$ is open in $H^s(M,M)$, it holds that $T_\varphi \mathrm{Diff}^s(M) = T_\varphi H^s(M,M)$. In particular, $T_{\mathrm{id}}\mathrm{Diff}^s(M) = \mathfrak{X}^s(M)$, i.e., the vector fields on M of Sobolev type H^s.

If $\psi \in \mathrm{Diff}^s(M)$, then right multiplication $\mathrm{Diff}^s(M) \ni \varphi \mapsto \varphi\circ\psi \in \mathrm{Diff}^s(M)$ is smooth. However, $\mathrm{Diff}^s(M)$ is *not* a Banach Lie group, because left multiplication is *not* smooth. Instead, $\mathrm{Diff}^s(M)$ is a topological group, i.e., the group operations are continuous. For details, see [7, § 2].

Let us now introduce the lifted inertia operator, given by $\mathcal{A}^\sharp := \sharp \circ \mathcal{A}$, with $\mathcal{A}$ defined by equation (1). We then have the following.

Lemma 3.1. *$\mathcal{A}^\sharp$ is a smooth isomorphism $\mathfrak{X}^s(M) \to \mathfrak{X}^{s-2}(M)$.*

Proof. Let $R : \Omega^{1,s}(M) \to \mathcal{H}^1(M)$ and $P_{\text{ex}} : \Omega^{1,s}(M) \to \delta\Omega^{1,s+1}(M)$ be the Hodge projections onto $\mathcal{H}^1(M)$ and $\delta\Omega^{1,s+1}(M)$ respectively. These are smooth mappings since the Hodge decomposition of $\Omega^{1,s}(M)$ is smooth (see [7]). Thus, since the musical isomorphism is smooth, the operator $Z : u \mapsto (Ru^\flat, P_{\text{ex}}u^\flat, \operatorname{div}(u))$ is a smooth mapping $\mathfrak{X}^s(M) \to \mathcal{H}^1(M) \times \delta\Omega^{1,s+1}(M) \times \mathcal{F}^{s-1}(M)$. For every s, it is in fact an isomorphism, so it has a smooth inverse which is given by $(h, \sigma, \rho) \mapsto h^\sharp + \sigma^\sharp + \operatorname{grad}(\Delta^{-1}(\rho))$. (Notice that $\Delta^{-1} : \mathcal{F}^{s-1}(M) \to \mathcal{F}^{s+1}(M)$.)

From the definition (1) of $\mathcal{A}$ it follows that

$$\mathcal{A}^\sharp = Z^{-1} \circ (\mathrm{id}, (1-\gamma)\mathrm{id} - \alpha\Delta, -\beta\Delta) \circ Z$$

This is an isomorphism $\mathfrak{X}^s(M) \to \mathfrak{X}^{s-2}(M)$ since $((1-\gamma)\mathrm{id} - \alpha\Delta) : \delta\Omega^{1,s+1}(M) \to \delta\Omega^{1,s-1}(M)$ and $\Delta : \mathcal{F}^{s-1}(M) \to \mathcal{F}^{s-3}(M)$ are isomorphisms (see [32]). The inverse is given by

$$(\mathcal{A}^\sharp)^{-1} = Z^{-1} \circ \big(\mathrm{id}, \big((1-\gamma)\mathrm{id} - \alpha\Delta\big)^{-1}, -\frac{1}{\beta}\Delta^{-1}\big) \circ Z.$$

This concludes the result. $\square$

From the definition of u it follows that $u(\varphi(x)) = \dot\varphi(x)$ for $x \in M$. By differentiating this with respect to t, we obtain

$$\frac{\mathrm{d}}{\mathrm{d}t}\Big(u\big(\varphi(x)\big)\Big) = \ddot\varphi(x) \in T^2_{(\varphi(x),\dot\varphi(x))}M. \tag{6}$$

The Levi–Civita connection ∇, induced by the Riemannian metric g on M, defines a diffeomorphism between the second tangent bundle T^2M and the Whitney sum $TM \oplus TM$ by $(c, \dot c, \ddot c) \mapsto (c, \dot c, \nabla_{\dot c}\dot c)$. By pointwise operations, this identifies the second tangent bundle $T^2\mathrm{Diff}^s(M)$ with the Whitney sum $T\mathrm{Diff}^s(M) \oplus T\mathrm{Diff}^s(M)$. By the ω–lemma (see e.g. [7]) the identification is smooth. Using this identification, and the fact that $u = \dot\varphi \circ \varphi^{-1}$, we can express equation (6) as

$$\dot u + \nabla_u u = \Big(\frac{D}{\mathrm{d}t}\dot\varphi\Big) \circ \varphi^{-1},$$

where $\frac{D}{\mathrm{d}t}\dot\varphi(x) := \nabla_{\dot\varphi(x)}\dot\varphi(x)$ is the co-variant derivative along the path itself. We can now write equation (2a) as

$$\mathcal{A}^\sharp\Big(\big(\frac{D}{\mathrm{d}t}\frac{\mathrm{d}\varphi}{\mathrm{d}t}\big) \circ \varphi^{-1}\Big) = -(\mathcal{L}_u\mathcal{A}u)^\sharp - (\mathcal{A}^\sharp u)\operatorname{div}(u) + \mathcal{A}^\sharp\nabla_u u =: F(u). \tag{7}$$

The approach is to show that this defines a smooth spray on $\mathrm{Diff}^s(M)$, i.e., a smooth vector field

$$\tilde S : T\mathrm{Diff}^s(M) \to T^2\mathrm{Diff}^s(M) \simeq T\mathrm{Diff}^s(M) \oplus T\mathrm{Diff}^s(M).$$

Let $R_\psi : \mathrm{Diff}^s(M) \to \mathrm{Diff}^s(M)$ denote composition with $\psi \in \mathrm{Diff}^s(M)$ from the right, i.e., $R_\psi(\varphi) = \varphi \circ \psi$. As already mentioned, this is a smooth mapping, so the corresponding tangent mapping TR_ψ, given by $T_\varphi\mathrm{Diff}^s(M) \ni v \mapsto v \circ \psi \in T_{\varphi\circ\psi}\mathrm{Diff}^s(M)$,

is also smooth. Let $T\mathrm{Diff}^{s-2}(M)\!\upharpoonright\!\mathrm{Diff}^s(M)$ denote the restriction of the tangent bundle $T\mathrm{Diff}^{s-2}(M)$ to the base $\mathrm{Diff}^s(M)$. This is a smooth Banach vector bundle (see [7, Appendix A]). If $B : \mathfrak{X}^s(M) \to \mathfrak{X}^{s-2}(M)$ then we denote by $\tilde{B}$ the bundle mapping $T\mathrm{Diff}^s(M) \to T\mathrm{Diff}^{s-2}(M)\!\upharpoonright\!\mathrm{Diff}^s(M)$ given by

$$\tilde{B}(\varphi, \dot{\varphi}) \to (\varphi, \tilde{B}_\varphi(\dot{\varphi})), \quad \tilde{B}_\varphi(\dot{\varphi}) := TR_\varphi \circ B \circ TR_{\varphi^{-1}}.$$

If B is smooth, then the mapping $\tilde{B}_\varphi : T_\varphi \mathrm{Diff}^s \to T_\varphi \mathrm{Diff}^{s-2}(M)$ is smooth for fixed $\varphi \in \mathrm{Diff}^s(M)$. However, in general, even if B is smooth, the mapping $\tilde{B}$ need not be smooth. This is because the operation $\varphi \mapsto \varphi^{-1}$ is not smooth. However, the following key lemmas resolve the situation in our specific case.

Lemma 3.2. *The mapping*

$$\tilde{\mathcal{A}}^\sharp : T\mathrm{Diff}^s(M) \to T\mathrm{Diff}^{s-2}(M)\!\upharpoonright\!\mathrm{Diff}^s(M)$$

is a smooth vector bundle isomorphism.

Proof. We have

$$\tilde{\mathcal{A}}^\sharp = \tilde{R}^\sharp + (1-\gamma)\tilde{P}^\sharp_{\mathrm{ex}} + \tilde{W}$$

where $P^\sharp_{\mathrm{ex}} = \sharp \circ P_{\mathrm{ex}} \circ \flat$, $R^\sharp = \sharp \circ R \circ \flat$ are the lifted Hodge projections and $W = \alpha\, \sharp \circ \delta \circ \mathrm{d} \circ \flat - \beta\ \mathrm{grad} \circ \mathrm{div}$. $\tilde{P}^\sharp_{\mathrm{ex}}$ and $\tilde{R}^\sharp$ are smooth bundle maps $T\mathrm{Diff}^s(M) \to T\mathrm{Diff}^s(M)$, see [7, Appendix A, Lemmas 2,3,6]. Thus, $\tilde{P}^\sharp_{\mathrm{ex}}$ and $\tilde{R}^\sharp$ are also smooth as mappings $T\mathrm{Diff}^s(M) \to T\mathrm{Diff}^{s-2}(M)\!\upharpoonright\!\mathrm{Diff}^s(M)$. That $\tilde{W}$ is smooth follows from [7, Appendix A, Lemma 2].

In a local chart in a neighbourhood of $\varphi \in \mathrm{Diff}^s(M)$, the derivative of $\tilde{\mathcal{A}}^\sharp$ at $(\varphi, \dot{\varphi})$ is a smooth linear mapping of the form

$$\begin{pmatrix} \mathrm{id} & 0 \\ * & \tilde{\mathcal{A}}^\sharp_\varphi \end{pmatrix} : T_\varphi \mathrm{Diff}^s(M) \times T_\varphi \mathrm{Diff}^s(M) \to T_\varphi \mathrm{Diff}^s(M) \times T_\varphi \mathrm{Diff}^{s-2}(M)$$

It follows from Lemma 3.1 that $\tilde{\mathcal{A}}^\sharp_\varphi$ is a linear isomorphism, with smooth inverse given by $\widetilde{(\mathcal{A}^\sharp)^{-1}_\varphi}$. The result now follows from the inverse function theorem for Banach manifolds. □

Lemma 3.3. *Let $B : \mathfrak{X}^s(M) \to \mathfrak{X}^{s-k}(M)$ be a smooth linear differential operator of order k. If $s > n/2 + k$, then the mapping*

$$u \longmapsto B\nabla_u u - \nabla_u B u = [B, \nabla_u]u$$

is a smooth non-linear differential operator $\mathfrak{X}^s(M) \to \mathfrak{X}^{s-k}(M)$.

Proof. If f and g are a scalar differential operators of order k and l respectively, then $[f, g]$ is a scalar differential operator of order $k + l - 1$, since the order $k + l$ differential terms in the commutator cancel each other. In general, this is not true for vector valued

differential operators. However, for a fixed v, the linear operator $u \mapsto \nabla_v u$ is given in components by

$$\nabla_v u = \Big(v^i u^j \Gamma^k_{ij} + v^i \frac{\partial u^k}{\partial x^i}\Big)\mathbf{e}_k,$$

so the part of ∇_v that is differentiating is acting diagonally on the elements of u. We write $\nabla_v u = Gu + f(u^i)\mathbf{e}_i$, where $G : \mathfrak{X}^s(M) \to \mathfrak{X}^s(M)$ is tensorial and f is a scalar differential operator of order 1. If $B = (b^i_j)$, so that $B(u_1\partial_1 + \ldots + u_n\partial_n) = b^i_j(u^j)\mathbf{e}_i$, then

$$[B, \nabla_v]u = [B, G]u + \big([f, b^i_j]u^j\big)\mathbf{e}_i.$$

Since G is tensorial, $[B, G]$ is a differential operator of the same order as B, that is k. Since f and b_{ij} are scalar differential operators of order 1 and k, it holds that $[f, b_{ij}]$ is of order $k + 1 - 1 = k$. Since $\nabla_v Bu$ differentiates v zero times, and $B\nabla_v u$ differentiates v at most k times, it is now clear that the total operation $u \mapsto [B, \nabla_u u]$ differentiates u at most k times. This finishes the proof. □

Lemma 3.4. *Let $B : \mathfrak{X}^s(M) \to \mathfrak{X}^{s-k}(M)$ be a smooth linear differential operator of order k. If $s > n/2 + k$, then the mapping*

$$\tilde{B} : T\mathrm{Diff}^s(M) \to T\mathrm{Diff}^{s-k}(M) \restriction \mathrm{Diff}^s(M)$$

is a smooth bundle map.

Proof. Assume first that B is of order 1. Then, locally, $\tilde{B}(\varphi, \dot{\varphi})$ is constructed by rational combinations of $\varphi^i, \dot{\varphi}^i, \frac{\mathrm{d}\varphi^i}{\mathrm{d}x^j}, \frac{\mathrm{d}\dot{\varphi}^i}{\mathrm{d}x^j}$. Smoothness then follows since pointwise multiplications are smooth operations (see [7, Appendix A, Lemma 2]). We can now, at least locally, decompose B into the composition of first order operators, so that $B = B_1 \cdots B_k$. It then holds that $\tilde{B} = \tilde{B}_1 \cdots \tilde{B}_k$, and by the argument above, each $\tilde{B}_i$ is smooth and drops differentiability by 1. This finishes the proof. □

Remark 3.5. Notice that if $Q : \mathfrak{X}^s(M) \times \mathfrak{X}^{s-l}(M) \to \mathfrak{X}^{s-k}$ is a bilinear differential operator, of order $k \geq 0$ in its first argument and order $k - l \geq 0$ in its second argument, then Lemma 3.4 implies that

$$\tilde{Q} : T\mathrm{Diff}^s(M) \times T\mathrm{Diff}^{s-l}(M) \restriction \mathrm{Diff}^s(M) \to T\mathrm{Diff}^{s-k}(M) \restriction \mathrm{Diff}^s(M)$$

is smooth whenever $s > n/2 + k$.

Lemma 3.6. *If $s > n/2 + 2$, then the mapping*

$$\tilde{F} : T\mathrm{Diff}^s(M) \to T\mathrm{Diff}^{s-2}(M) \restriction \mathrm{Diff}^s(M)$$

is a smooth bundle map.

Proof. For any $v \in \mathfrak{X}(M)$ we have $(\mathcal{L}_u v^\flat)^\sharp = \mathcal{L}_u v + 2\mathrm{Def}(u)v$, where $\mathrm{Def}(u)$ is the type $(1,1)$ tensor defined by $\frac{1}{2}(\mathcal{L}_u \mathsf{g})(v,\cdot) = \mathsf{g}(\mathrm{Def}(u)v,\cdot)$ (see e.g. [32, § 2.3]). Thus

$$\begin{aligned}
F(u) &= -(\mathcal{L}_u \mathcal{A}u)^\sharp - (\mathcal{A}^\sharp u)\,\mathrm{div}(u) + \mathcal{A}^\sharp \nabla_u u \\
&= -\mathcal{L}_u \mathcal{A}^\sharp u - 2\mathrm{Def}(u)\mathcal{A}^\sharp u - (\mathcal{A}^\sharp u)\,\mathrm{div}(u) + \mathcal{A}^\sharp \nabla_u u \\
&= -\nabla_u \mathcal{A}^\sharp u + \nabla_{\mathcal{A}^\sharp u} u - 2\mathrm{Def}(u)\mathcal{A}^\sharp u - (\mathcal{A}^\sharp u)\,\mathrm{div}(u) + \mathcal{A}^\sharp \nabla_u u \\
&= \underbrace{(\mathcal{A}^\sharp \nabla_u - \nabla_u \mathcal{A}^\sharp)}_{[\mathcal{A}^\sharp,\nabla_u]} u + \nabla_{\mathcal{A}^\sharp u} u - 2\mathrm{Def}(u)\mathcal{A}^\sharp u - (\mathcal{A}^\sharp u)\,\mathrm{div}(u),
\end{aligned}$$

where we have used that $\mathcal{L}_u v = \nabla_u v - \nabla_v u$.

Let $Q : \mathfrak{X}^s(M) \times \mathfrak{X}^{s-2}(M) \to \mathfrak{X}^{s-2}(M)$ be the bilinear mapping

$$Q(u,v) := \nabla_v u - 2\mathrm{Def}(u)v - v\,\mathrm{div}(u).$$

Notice that Q is tensorial in v and of order 1 in u. If $s > n/2 + 2$ then Q is smooth. Write $\mathcal{A}^\sharp = P + W$, where $P = R^\sharp + (1-\gamma)P_{\mathrm{ex}}^\sharp$ and W is a linear differential operator of order 2 as above. We now have

$$F(u) = [\mathcal{A}^\sharp, \nabla_u]u + Q(u, \mathcal{A}^\sharp u) = [P,\nabla_u]u + [W,\nabla_u]u + Q(u,Pu) + Q(u,Wu).$$

The approach is to show that each of these terms are of maximal order 2 and smooth under conjugation with right translation.

For the first term, we have

$$\widetilde{[P,\nabla_{(\cdot)}]} = \tilde{P} \circ \widetilde{\nabla_{(\cdot)}} - \widetilde{\nabla_{(\cdot)}} \circ \tilde{P}$$

We already know that $\tilde{P} : T\mathrm{Diff}^s(M) \to T\mathrm{Diff}^s(M)$ is smooth. From Lemma 3.4 and Remark 3.5 it follows that $\widetilde{\nabla_{(\cdot)}} : T\mathrm{Diff}^s(M) \to T\mathrm{Diff}^{s-1}(M) \restriction \mathrm{Diff}^s(M)$ is smooth.

For the second term, $(u,v) \mapsto [W,\nabla_v]u$ is a bilinear differential operator. From Lemma 3.3 it follows that it is of order 2 (since W is of order 2). From Lemma 3.4 and Remark 3.5 it then follows that $\widetilde{[W,\nabla_{(\cdot)}]}$ is smooth.

For the third term, it follows from Lemma 3.4 and Remark 3.5 that $\tilde{Q}$ is smooth of order 1. Since $\tilde{P}$ is smooth of order 0, we get that $\widetilde{Q(\cdot, P\,\cdot)}$ is smooth of order 1.

For the fourth term, $(u,v) \mapsto Q(u,Wv)$ is a bilinear differential operator of order 1 and 2 in its arguments. It then follows from Lemma 3.4 and Remark 3.5 that $\widetilde{Q(\cdot, W\,\cdot)}$ is smooth of order 2.

Altogether, we now have that $\tilde{F} : T\mathrm{Diff}^s(M) \to T\mathrm{Diff}^{s-2}(M) \restriction \mathrm{Diff}^s(M)$ is smooth, which finishes the proof. □

Equation (7) can be written

$$\tilde{\mathcal{A}}^\sharp(\varphi, \frac{D}{\mathrm{d}t}\dot{\varphi}) = \tilde{F}(\varphi, \dot{\varphi}). \tag{8}$$

We now obtain the main result in this section.

Theorem 3.7. *If $s > n/2 + 2$, then the geodesic spray*

$$\tilde{S} : T\mathrm{Diff}^s(M) \ni (\varphi, \dot{\varphi}) \mapsto \Big(\varphi, \dot{\varphi}, \big((\tilde{\mathcal{A}}^\sharp)^{-1} \circ \tilde{F}\big)(\varphi, \dot{\varphi})\Big) \in T\mathrm{Diff}^s(M) \oplus T\mathrm{Diff}^s(M)$$

corresponding to the α-β-γ metric (5) on $\mathrm{Diff}^s(M)$ is smooth.

Proof. Follows from Lemma 3.2 and Lemma 3.6. □

In turn, this result implies that the geodesic equation is locally well posed, and that the solution depends smoothly on the initial data.

Corollary 3.8. *Under the same conditions as in Theorem 3.7, the Riemannian exponential*

$$\mathrm{Exp} : T\mathrm{Diff}^s(M) \to \mathrm{Diff}^s(M)$$

corresponding to the α-β-γ metric (5) on $\mathrm{Diff}^s(M)$ is smooth. Further, if $\varphi \in \mathrm{Diff}^s(M)$ then

$$\mathrm{Exp}_\varphi : T_\varphi\mathrm{Diff}^s(M) \to \mathrm{Diff}^s(M)$$

is a local diffeomorphism from a neighbourhood of 0 to a neighbourhood of φ.

Proof. Follows from standard results about smooth sprays on Banach manifolds [22]. □

4 Descending metrics and the space of densities

Let $\pi : E \to B$ be a smooth fibre bundle. In this section we characterise pairs of Riemannian metrics on E and B for which the projection π is a Riemannian submersion. We do this in three steps, by introducing more and more structure to the fibre bundle:

1. First, the plain case $\pi : E \to B$. A basic characterisation of all the descending metrics is given.

2. Second, the case when $\pi : E \to B$ is a principle H–bundle. This allows us to characterise descending metrics in terms of H–invariance.

3. Third, the case when $E = G$, where G is a Lie group, and B is a G–homogeneous space, i.e., there is a transitive Lie group action of G on B. Then the projection $\pi_b : g \mapsto b \cdot g$, for any fixed element $b \in B$, defines a principle G_b–bundle, where G_b is the isotropy group of b. This structure allows us to consider metrics on G which are both right invariant and descending.

The main example is $G = \mathrm{Diff}(M)$ and $B = \mathrm{Dens}(M)$, i.e, the space of smooth densities on M (see below). The main result is that the α-β-γ metric (5) on $\mathrm{Diff}(M)$ descend to the right invariant canonical L^2 metric on $\mathrm{Dens}(M)$ (the Fisher metric).

Let us begin with the plain case. The kernel of the derivative of the projection map π defines the vertical distribution $\mathcal{V}$ on E, i.e., for each $x \in E$

$$\mathcal{V}_x = \{v \in T_xE\,;\, T_x\pi \cdot v = 0\}.$$

If g_E is a Riemannian metric on E, then we can also define the horizontal distribution $\mathcal{H} = \mathcal{V}^\perp$ as the orthogonal complement of $\mathcal{V}$ with respect to g_E.

Definition 4.1. A Riemannian metric g_E on E is called *descending* if there exists a Riemannian metric g_B on B such that

$$\mathsf{g}_E(u,v) = \pi^*\mathsf{g}_B(u,v) \quad \forall\, u,v \in \mathcal{H}.$$

Thus, a metric g_E on E is descending if and only if there exists a metric g_B on B such that π is a Riemannian submersion, i.e., $T\pi : TE \to TB$ preserves the length of horizontal vectors.

If g_E is a descending metric, then the metric g_B is unique. This follows since

$$T_x\pi : \mathcal{H}_x \to T_{\pi(x)}B$$

is an isomorphism for each $x \in E$.

We now show how to construct descending metrics. Let g_B be any Riemannian metric on B. Then we can lift g_B to a positive semi-definite bilinear form $\pi^*\mathsf{g}_B$ on E. Next, let h be another positive semi-definite bilinear form on E such that $\ker(\mathsf{h}) \cap \mathcal{V} = \{0\}$ and the co-dimension of $\ker(\mathsf{h})$ is equal to the dimension of $\mathcal{V}$. Then

$$\mathsf{g}_E = \pi^*\mathsf{g}_B + \mathsf{h} \tag{9}$$

is a descending Riemannian metric on E. Notice that $\ker(\pi^*\mathsf{g}_B) = \mathcal{V}$ and $\ker(\mathsf{h}) = \mathcal{H}$. Thus, $\mathsf{g}_E(u,v) = \pi^*\mathsf{g}_B(u,v)$ for all $u,v \in \mathcal{H}$, so g_E is indeed descending. Notice also that the horizontal distribution is independent of the choice of g_B.

The form (9) characterises all descending metrics. Indeed, if g_E is a descending metric, let g_B be the corresponding metric on B and let $P : TE \to \mathcal{V}$ be the orthogonal projection onto $\mathcal{V}$ with respect to g_E. Then g_E is of the form (9) with $\mathsf{h}(u,v) := \mathsf{g}_E(u, Pv)$.

Consider now the second case. That is, let H be a Lie group and consider the case when $\pi : E \to B$ is a principle H–bundle, with a left action $L_h : E \to E$ for $h \in H$. Being a principle bundle, the fibres are parameterised by H, so $\pi \circ L_h = \pi$ and if $\pi(x) = \pi(y)$ then there exists a unique $h \in H$ such that $y = L_h(x)$. Thus, if g_B is a Riemannian metric on B, then $\pi^*\mathsf{g}_B = (\pi \circ L_h)^*\mathsf{g}_B = L_h^*\pi^*\mathsf{g}_B$. It follows that if g_E is a descending metric, then

$$L_h^*\mathsf{g}_E(u,v) = \mathsf{g}_E(u,v) \quad \forall u,v \in \mathcal{H}. \tag{10}$$

The converse is also true.

Proposition 4.2. *Let g_E be a Riemannian metric on E. Then g_E is descending if and only if it fulfils (10).*

Proof. We have already shown $\Rightarrow$ so $\Leftarrow$ remains. Assuming (10), define g_B in the following way. For $\bar{u}, \bar{v} \in T_{\pi(x)}B$, take any point $y \in \pi^{-1}(\{x\})$. The linear map $T_y\pi : \mathcal{H}_y \to T_{\pi(x)}B$ is an isomorphism, so we get $u, v \in \mathcal{H}_y$ by $u = T_y\pi^{-1} \cdot \bar{u}$ and $v = T_y\pi^{-1} \cdot \bar{v}$. Now, define g_B by

$$\mathsf{g}_B(\bar{u}, \bar{v}) := \mathsf{g}_E(u,v).$$

This is a well define metric on g_E, i.e., it is independent on which $y \in \pi^{-1}(\{x\})$ we use. Indeed, for another $y' \in \pi^{-1}(\{x\})$ we get $u', v' \in \mathcal{H}_{y'}$ as above. Also, $y' = L_h(y)$ for

some $h \in H$. From (10) it then follows that $\mathsf{g}_E(u,v) = \mathsf{g}_E(u',v')$, so g_B is well defined. By construction, $\mathsf{g}_E(u,v) = \pi^*\mathsf{g}_B(u,v)$ for all $u,v \in \mathcal{H}$, so g_E is indeed a descending metric. □

We now specialise even further to the third case. That is, let G be a Lie group with identity e. Denote by L_g and R_g respectively the left and right action of $g \in G$ on G. Assume that G has a right transitive action $\bar{R}_g$ on a manifold B. (B is then called a G–homogeneous space.) If $b \in B$, then $G_b = \{g \in G; \bar{R}_g(b) = b\}$ denotes the isotropy Lie subgroup of G. For every $b \in B$ we then have a principle G_b–bundle $\pi_b : G \to B$, where $\pi_b(g) = \bar{R}_g(b)$. Notice that this structure implies that B is diffeomorphic to the homogeneous space $G_b \backslash G$ of right co-sets. The map π_b, which is well defined on $G_b\backslash G$, provides a diffeomorphism. (If $b, b' \in B$ we also have that G_b and $G_{b'}$ are conjugate subgroups, i.e., there exists a $g \in G$ such that $gG_bg^{-1} = G_{b'}$.)

We are interested in Riemannian metrics g_G on G which are right invariant, i.e., for which

$$\mathsf{g}_G(u,v) = \mathsf{g}_G(TR_g \cdot u, TR_g \cdot u)$$

or equivalently $R_g^*\mathsf{g}_G = \mathsf{g}_G$. Notice that right invariance *does not* imply that the metric is descending. Indeed, in order for a right invariant metric g_G to be descending with respect to π_b, it follows from Proposition 4.2 that

$$L_h^* R_g^* \mathsf{g}_G(u,v) = \mathsf{g}_G(u,v) \quad \forall u,v \in \mathcal{H}^b, \quad \forall g \in G, \quad \forall h \in G_b$$

where $\mathcal{H}^b$ denotes the horizontal distribution. Since g_G is right invariant and since the right action descends to $G_b\backslash G$, i.e., R_g maps fibres to fibres, it is enough to check the condition for $g = h^{-1}$ and for vectors $u,v \in \mathcal{H}_e^b = \mathfrak{g}_b^\perp$, where $\mathfrak{g}_b$ is the Lie algebra of G_b. Indeed, the following result is given in [19].

Proposition 4.3. *Let g_G be a right invariant metric on G. Then g_G is descending (with respect to π_b) if and only if*

$$\mathsf{g}_G(\mathrm{ad}_\xi(u), v) + \mathsf{g}_G(u, \mathrm{ad}_\xi(v)) = 0 \quad \forall\, u,v \in \mathfrak{g}_b^\perp,\ \xi \in \mathfrak{g}_b.$$

Consider now the reverse question, i.e., if $\mathsf{g}_G = \pi^*\mathsf{g}_B + \mathsf{h}$ is a descending metric, when is it right invariant? Since right invariance means that $\mathsf{g}_G = R_g^*\mathsf{g}_G$ it must hold that $R_g^*\pi^*\mathsf{g}_B = \pi^*\mathsf{g}_B$ and $R_g^*\mathsf{h} = \mathsf{h}$. Also, since

$$R_g^*\pi^*\mathsf{g}_B = (\pi \circ R_g)^*\mathsf{g}_B = (\bar{R}_g \circ \pi)^*\mathsf{g}_B = \pi^*\bar{R}_g^*\mathsf{g}_B$$

we obtain the following result.

Proposition 4.4. *Let $\mathsf{g}_G = \pi^*\mathsf{g}_B + \mathsf{h}$ be a descending Riemannian metric on G. Then g_G is right invariant if and only if both g_B and h are right invariant, i.e.,*

$$\bar{R}_g^*\mathsf{g}_B = \mathsf{g}_B \qquad \text{and} \qquad R_g^*\mathsf{h} = \mathsf{h}$$

for all $g \in G$.

We now investigate what the geometric concepts investigated in this section implies in the case $G = \mathrm{Diff}(M)$.

First, we introduce the manifold of smooth probability densities on M, which takes the rôle of the manifold B above. It is given by

$$\mathrm{Dens}(M) = \left\{ \nu \in \Omega^n(M)\,;\, \nu > 0, \int_M \nu = 1 \right\}.$$

The tangent spaces of $\mathrm{Dens}(M)$ are $T_\nu \mathrm{Dens}(M) = \Omega^n_0(M) := \{a \in \Omega^n(M); \int_M a = 0\}$.

$\mathrm{Diff}(M)$ acts on $\mathrm{Dens}(M)$ from the right by pullback $\bar{R}_\varphi(\nu) = \varphi^*\nu$. The corresponding lifted action is again given by pullback, i.e., $T\bar{R}_\varphi(\nu, a) = (\varphi^*\nu, \varphi^* a)$.

Consider now the volume form $\mathrm{vol} \in \mathrm{Dens}(M)$, corresponding to the Riemannian structure on M. Using the action $\bar{R}_\varphi$ and this volume form vol, we define the projection map $\pi_{\mathrm{vol}} : \mathrm{Diff}(M) \to \mathrm{Dens}(M)$ by $\pi_{\mathrm{vol}}(\varphi) = \bar{R}_\varphi(\mathrm{vol}) = \varphi^*\mathrm{vol}$. This map is a submersion, since the action is transitive, which is proved by Moser [28]. Furthermore, the corresponding isotropy group is given by $\mathrm{Diff}_{\mathrm{vol}}(M)$, i.e., if $\psi \in \mathrm{Diff}_{\mathrm{vol}}(M)$ then $\pi_{\mathrm{vol}}(\psi \circ \varphi) = \pi_{\mathrm{vol}}(\varphi)$. Accordingly, with $\mathrm{Diff}_{\mathrm{vol}}(M)$ acting on $\mathrm{Diff}(M)$ from the left, we have the principle bundle structure

$$\mathrm{Diff}_{\mathrm{vol}}(M) \hookrightarrow \mathrm{Diff}(M) \xrightarrow{\pi_{\mathrm{vol}}} \mathrm{Dens}(M). \tag{11}$$

The vertical distribution $\mathcal{V}$ of this bundle structure is given by vectors in $T\mathrm{Diff}(M)$ which, right translated to $T_{\mathrm{id}}\mathrm{Diff}(M) = \mathfrak{X}(M)$, are divergence free vector fields. That is,

$$\mathcal{V}_\varphi = \{v \in T_\varphi \mathrm{Diff}(M)\,;\, v \circ \varphi^{-1} \in \mathfrak{X}_{\mathrm{vol}}(M)\}.$$

In reference to the abstract formulation above: $G = \mathrm{Diff}(M)$, $B = \mathrm{Dens}(M)$, and $G_b = \mathrm{Diff}_{\mathrm{vol}}(M)$. In particular, $\mathrm{Diff}_{\mathrm{vol}}(M)\backslash\mathrm{Diff}(M) \simeq \mathrm{Dens}(M)$.

Remark 4.5. More specifically, it holds that $\mathrm{Diff}^s_{\mathrm{vol}}(M)\backslash\mathrm{Diff}^s(M) \simeq \mathrm{Dens}^{s-1}(M)$ if $s > n/2+1$. Also, in this setting the projection $\pi_{\mathrm{vol}} : \mathrm{Diff}^s(M) \to \mathrm{Dens}^{s-1}(M)$ is smooth. These results are then used to show that $\pi_{\mathrm{vol}} : \mathrm{Diff}(M) \to \mathrm{Dens}(M)$ is smooth with respect to the ILH topology. Notice, however, that the principle bundle structure $\pi_{\mathrm{vol}} : \mathrm{Diff}^s(M) \to \mathrm{Diff}^s(M)/\mathrm{Diff}^s_{\mathrm{vol}}(M)$ is only C^0, since the left action of $\mathrm{Diff}^s(M)$ on itself is only continuous. See Ebin and Marsden [7, § 5] for details.

By using the Nash–Moser inverse function theorem, Hamilton [13, §III.2.5] also showed directly that $\pi_{\mathrm{vol}} : \mathrm{Diff}(M) \to \mathrm{Dens}(M)$ is a smooth principle $\mathrm{Diff}_{\mathrm{vol}}(M)$–bundle with respect to a Fréchet topology.

Remark 4.6. As already mentioned in the abstract setting, the choice of reference element $\mathrm{vol} \in \mathrm{Dens}(M)$ has no canonical meaning. It simply specifies which point in $\mathrm{Dens}(M)$ we consider to be the "identity density". Indeed, if $\nu \in \mathrm{Dens}(M)$ is another density, then $\mathrm{Diff}_{\mathrm{vol}}(M)$ and $\mathrm{Diff}_\nu(M)$ are conjugate subgroups, i.e., there exists a $\psi \in \mathrm{Diff}(M)$ such that $\mathrm{Diff}_\nu(M) = \psi \circ \mathrm{Diff}_{\mathrm{vol}}(M) \circ \psi^{-1}$.

Consider now the question of right invariant and descending metrics on $\mathrm{Diff}(M)$. First, there is a natural L^2 metric on $\mathrm{Dens}(M)$, given by

$$\langle\!\langle a, b \rangle\!\rangle_\nu = \int_M \frac{\mathrm{d}a}{\mathrm{d}\nu}\frac{\mathrm{d}b}{\mathrm{d}\nu}\,\nu, \qquad a, b \in \Omega^n_0(M), \tag{12}$$

where $\mathrm{d}a/\mathrm{d}\nu, \mathrm{d}b/\mathrm{d}\nu \in \Omega^0(M)$ are the Radon–Nikodym derivatives of a and b with respect to ν. This metric is called the Fisher metric. As mentioned already, it is fundamental in the theory of information geometry. In addition, it is used in statistical mechanics, for measuring "thermodynamic length" (see e.g. [6, 10]), and in quantum mechanics (see e.g. [9]). Notice that the Fisher metric is canonical, in the sense that it is independent of the Riemannian structure on M.

Remark 4.7. One can also write the Fisher metric (12) as

$$\langle\!\langle a, b \rangle\!\rangle_\nu = \int_M (\star_\nu a) b,$$

where $\star_\nu : \Omega^n(M) \to \Omega^0(M)$ is the Hodge star on n–forms corresponding to ν.

Proposition 4.8. *The Fisher metric* (12) *on* $\mathrm{Dens}(M)$ *is invariant with respect to the action* $\bar{R}_\varphi$.

Proof.

$$\langle\!\langle \varphi^* a, \varphi^* b \rangle\!\rangle_{\varphi^*\nu} = \int_M (\star_{\varphi^*\nu}\varphi^* a)\varphi^* b = \int_M \varphi^*\big((\star_\nu a) b\big) = \langle\!\langle a, b \rangle\!\rangle_\nu.$$

□

We now come to the main result of this section. It states that the α-β-γ metric on $\mathrm{Diff}(M)$ descends to the Fisher metric on $\mathrm{Dens}(M)$.

Theorem 4.9. *The α-β-γ metric* (5) *on* $\mathrm{Diff}(M)$ *descends to a metric on* $\mathrm{Dens}(M)$*, which, up to multiplication with β, is the Fisher metric* (12)*.*

Proof. First, since the inner product (4) on $\mathfrak{X}(M)$ corresponding to the α-β-γ metric preserves orthogonality with respect to the Helmholtz decomposition, it follows that the horizontal distribution is given by

$$\mathcal{H}_\varphi = \big\{ v \in T_\varphi \mathrm{Diff}(M)\,;\, v \circ \varphi^{-1} \in \mathrm{grad}(\mathcal{F}(M)) \big\},$$

i.e., vectors that, when translated to the identity, are given by gradient vector fields. Now, if $u \in \mathcal{V}_{\mathrm{id}} = \mathfrak{X}_{\mathrm{vol}}(M)$ and $f, g \in \mathcal{F}(M)$, then

$$\begin{aligned}
\langle \mathcal{L}_u \,\mathrm{grad}(f), \mathrm{grad}(g) \rangle_{\alpha\beta\gamma} &= \int_M \beta\, \delta(\mathcal{L}_u \,\mathrm{grad}(f))^\flat \delta \,\mathrm{grad}(g)^\flat \,\mathrm{vol} \\
&= \int_M \beta \,\mathrm{d}i_{\mathcal{L}_u \mathrm{grad}(f)} \mathrm{vol} \wedge \star \,\mathrm{d}i_{\mathrm{grad}(g)} \mathrm{vol} \\
&= -\int_M \beta \,\mathrm{d}i_{\mathrm{grad}(f)} \mathrm{vol} \wedge \star \,\mathrm{d}i_{\mathcal{L}_u \mathrm{grad}(g)} \mathrm{vol} \\
&= -\langle \mathrm{grad}(f), \mathcal{L}_u \,\mathrm{grad}(g) \rangle_{\alpha\beta\gamma},
\end{aligned}$$

where we have used that $\mathcal{L}_u\mathrm{vol} = 0$, and that $\mathcal{L}_u \star a = \star\mathcal{L}_u a$ for any $a \in \Omega^n(M)$, which also follows since $u \in \mathfrak{X}_{\mathrm{vol}}(M)$. From Proposition 4.3 it now follows that the α-β-γ metric is descending.

The tangent map $T\pi_{\mathrm{vol}}$ restricted to $T_{\mathrm{id}}\mathrm{Diff}(M) = \mathfrak{X}(M)$ is given by $u \mapsto \mathcal{L}_u\mathrm{vol}$. Now,

$$
\begin{aligned}
\langle \mathrm{grad}(f), \mathrm{grad}(g)\rangle_{\alpha\beta\gamma} &= \int_M \beta\, \mathrm{i}_{\mathrm{grad}(f)}\mathrm{vol} \wedge \star \mathrm{i}_{\mathrm{grad}(g)}\mathrm{vol} \\
&= \int_M \beta\, \mathcal{L}_{\mathrm{grad}(f)}\mathrm{vol} \wedge \star\mathcal{L}_{\mathrm{grad}(f)}\mathrm{vol} \\
&= \beta \langle\!\langle \mathcal{L}_{\mathrm{grad}(f)}\mathrm{vol}, \mathcal{L}_{\mathrm{grad}(g)}\mathrm{vol}\rangle\!\rangle_{\mathrm{vol}}.
\end{aligned}
$$

Therefore, the α-β-γ metric for horizontal vectors at the identity tangent space is given by the Fisher metric multiplied by β of the projection of the horizontal vectors to the tangent space $T_{\mathrm{vol}}\mathrm{Dens}(M)$. Since this holds at one tangent space, it follows from Proposition 4.8 that it holds at every tangent space (both the α-β-γ metric and the Fisher metric are right invariant). This concludes the proof. ☐

As a consequence of this result, we now obtain a geometric explanation of the observation in § 1.1, that solutions which are initially gradient vector fields remain gradients. This is a consequence of a general property on Riemannian submersions, proved by Hermann [14], which is that initially horizontal geodesics remain horizontal.

Remark 4.10. The "components" g_B and h for the α-β-γ metric are identified as follows:

$$
\langle u, v\rangle_{\alpha\beta\gamma} = \underbrace{\langle \bar{P}_\gamma u^\flat, \bar{P}_\gamma v^\flat\rangle_{L^2} + \alpha\langle \mathrm{d}u^\flat, \mathrm{d}v^\flat\rangle_{L^2}}_{\mathsf{h}} + \underbrace{\beta\langle \delta u^\flat, \delta v^\flat\rangle_{L^2}}_{\pi^*\mathsf{g}_B},
$$

with the same notation as in equation (4).

5 Optimal transport and polar factorisation

The field of optimal transport has a long history, going back to Monge [27] and Kantorovich [17, 16]. For an overview of the subject, see, e.g., the the monograph by Villani [33], or the lecture notes by Evans [8] or McCann [26], and references therein.

In this section we study the relation between optimal transport problems and metrics on $\mathrm{Diff}(M)$ descending to $\mathrm{Dens}(M)$. Typically, optimal transport problems are considered with minimal restrictions on the regularity of maps and densities. Our setting is restricted to the smooth case (or more precisely Sobolev H^s). The point of view is that of Otto [29, § 4], but with the difference that $\mathrm{Dens}(M)$ is identified with right instead of left co-sets (see § 4). We also discuss the correspondence between optimal control problems and polar factorisation. The main result is given in § 5.1, where we establish existence and uniqueness of the optimal information transport problem corresponding to the α-β-γ metric (5), and a matching polar factorisation result for $\mathrm{Diff}^s(M)$. In addition, as a finite dimensional analogue, we show in § 5.2 that the QR factorisation of square matrices can be seen as a polar factorisation result corresponding to optimal transport of inner products on $\mathbb{R}^n$.

From a geometric point of view, one can define optimal transport problems abstractly as follows. Let G be a Lie group with identity element e, which is acting transitively on a manifold B. Assume that G is equipped with a cost function $c : G \times G \to \mathbb{R}^+$. A simple geometric formulation of Monge's original optimal transport problem is:

$$\text{Given } b, b' \in B, \text{ find } k \in \{g \in G; g \cdot b = b'\} \text{ minimising } c(e, k). \tag{13}$$

We are interested in the case when the cost function is $c(\cdot,\cdot) = \mathrm{dist}^2_G(\cdot,\cdot)$, where dist_G is a geodesic distance corresponding to a Riemannian metric g_G on G (we assume that G is connected, so that (G, dist_G) is a metric space). In particular, consider the case when g_G is descending with respect to a fibre bundle structure $\pi : G \to B$. Then, loosely speaking, the optimal transport problem reduces to: (i) finding a shortest curve on B that connects b and b', and (ii) lifting that curve to a horizontal geodesic in G and take the endpoint as the solution to (13). The basic reason is that a shortest curve (or a sequence of curves attaining the infimum length in the limit) between e and the fibre $\pi^{-1}(\{b'\})$ must be horizontal. Indeed, we have the following result.

Lemma 5.1. *Let $\pi : G \to B$ be a Riemannian submersion, and let $\zeta : [0,1] \to G$ be an arbitrary curve. Then there exists a unique horizontal curve $\zeta_h : [0,1] \to G$ such that $\zeta_h(0) = \zeta(0)$ and $\pi \circ \zeta = \pi \circ \zeta_h$. The length of ζ_h is less than or equal the length of ζ, with equality if and only if ζ is horizontal.*

Proof. For each $t \in [0,1]$ there is a unique decomposition $\dot\zeta(t) = v(t) + h(t)$, where $v(t) \in \mathcal{V}_{\zeta(t)}$ and $h(t) \in \mathcal{H}_{\zeta(t)}$. Thus, we have the curves $t \mapsto v(t) \in \mathcal{V}$ and $t \mapsto h(t) \in \mathcal{H}$. By the projection π we also get a curve $\bar\zeta(t) = \pi(\zeta(t)) \in B$. This curve can be lifted to a horizontal curve as follows. Take any time-dependent vector field $\bar X_t$ on B for which $\bar\zeta$ is an integral curve, i.e., $\dot{\bar\zeta}(t) = \bar X_t(\bar\zeta)$. Now lift $\bar X_t$ to its corresponding horizontal section $X_t(g) = (T_g\pi)^{-1} \cdot \bar X_t(\pi(g))$. (We can do this since $T_g\pi : \mathcal{V}_g \to T_{\pi(g)}B$ is an isomorphism.) Next, let ζ_h be the unique integral curve of X_t with $\zeta_h(0) = \zeta(0)$. By construction it holds that $\pi(\zeta_h(t)) = \pi(\zeta(t))$. By construction we also have that $\mathsf{g}_G(\dot\zeta_h(t), \dot\zeta_h(t)) = \mathsf{g}_G(h(t), h(t))$. Thus, $\mathsf{g}_G(\dot\zeta_h(t), \dot\zeta_h(t)) \leq \mathsf{g}_G(\dot\zeta(t), \dot\zeta(t))$, with equality if and only if $\dot\zeta(t) \in \mathcal{H}$.

It remains to show that ζ_h is unique. Assume that $\zeta'_h : [0,1] \to G$ is another horizontal curve such that $\pi \circ \zeta'_h = \bar\zeta$ and $\zeta'_h(0) = \zeta(0)$. By differentiation with respect to t we obtain

$$T_{\zeta'_h(t)}\pi \cdot \dot\zeta'_h(t) = \dot{\bar\zeta}(t) = \bar X_t(\bar\zeta(t)) = \bar X_t(\pi(\zeta'_h(t))) = T_{\zeta'_h(t)}\pi \cdot X_t(\zeta'_h(t)).$$

Since ζ'_h is horizontal, it follows that ζ'_h is an integral curve of X_t. Since ζ'_h fulfils the same initial condition as ζ_h, it follows from uniqueness of integral curves that $\zeta'_h = \zeta_h$, which concludes the proof. □

Remark 5.2. Notice that the result in Lemma 5.1 holds in the case when G and B are Banach manifolds, with a smooth bundle structure $\pi : G \to B$. It is *not* necessary that G is a Banach Lie group, i.e., that the group operations are smooth. This is important for the main example in § 5.1.

For a cost function corresponding to a descending Riemannian metric, Lemma 5.1 shows that the optimal transport problem (13) reduces to a problem entirely on B, namely to find a shortest curve between two given elements $b'', b' \in B$. It is not always the case that this problem is easier to solve. However, if the geometry of the Riemannian manifold (B, g_B) is well understood, for example if any two elements in B can be connected by a minimal geodesic, then the problem simplifies significantly.

Also related to optimal transport problems is the concept of *polar factorisation*. Indeed, following Brenier [4], we introduce the *polar cone* as the subset of G given by

$$K = \{k \in G; \mathrm{dist}_G(e, k) \le \mathrm{dist}_G(h, k),\ \forall h \in G_b\}.$$

Put in words, the polar cone is given by the elements in G for which the closest point on the identity fibre is e. We have the following result.

Proposition 5.3. *Let g_G be a metric on G which descends to a right invariant metric g_B on B with respect to the fibre structure $\pi_b(g) = \bar{R}_g(b)$ for some fixed $b \in B$. Then the following statements are equivalent:*

1. *If $b' \in B$, then there exists a unique minimal geodesic from b to b'.*
2. *If $b'', b' \in B$, then there exists a unique minimal geodesic from b'' to b'.*
3. *There exists a unique solution to the optimal transport problem (13), and that solution is connected to e by a unique minimal geodesic.*
4. *Every $g \in G$ has a unique factorisation $g = hk$, with $h \in G_b$ and $k \in K$, and every $k \in K$ is connected to e by a unique minimal geodesic.*

Proof. 1 $\Rightarrow$ 2. Since g_G is right invariant it holds that if $\zeta : [0,1] \to G$ is a minimal geodesic, then so is $\bar{R}_g \circ \zeta$ for any $g \in G$. Since the action $\bar{R}$ is transitive, $b'' = \bar{R}_g(b)$ for some $g \in G$.

2 $\Rightarrow$ 3. Let $\bar{\zeta} : [0,1] \to B$ be the minimal geodesic from b to b'. Then, by Lemma 5.1, there is a unique corresponding horizontal geodesic $\zeta : [0,1] \to G$ with $\zeta(0) = e$ and $\pi_b(\zeta(t)) = \bar{\zeta}(t)$. There cannot be any curve from e to $\pi_b^{-1}(\{b'\})$ which is shorter than ζ, because then $\bar{\zeta}$ would not be a minimal geodesic. A curve from e to $\pi_b^{-1}(\{b'\})$ of the same length as ζ must be horizontal (follows from Lemma 5.1), and therefore equal to ζ (which also follows from Lemma 5.1). Thus, if $q \in \pi_b^{-1}(\{b'\}) \backslash \{\zeta(1)\}$ then $\mathrm{dist}_G(e, q) > \mathrm{dist}_G(e, \zeta(1))$, so $\zeta(1)$ is the unique solution to problem (13). Also, ζ is a unique minimal geodesic between e and $\zeta(1)$.

3 $\Rightarrow$ 4. Let k be the unique solution to (13) with $b' = \pi_b(g)$. Then k and g belong to the same fibre, so $g = hk$ for some unique element $h \in G_b$. There cannot be another such factorisation $g = h'k'$, because then k would not be a unique solution to (13). Now take any $k \in K$. Then k is the unique solution to (13) with $b' = \pi_b(k)$, and that solution is connected to e by a unique minimal geodesic. Thus, any $k \in K$ is connected to e by a unique minimal geodesic.

4 $\Rightarrow$ 1. Since the action $\bar{R}$ is transitive, we can find a $g \in G$ such that, for any $b' \in B$, it holds that $b' = \bar{R}_g(b) = \pi_b(g)$. Let $g = hk$ be the unique factorisation, and

let $\zeta : [0,1] \to G$ be the unique minimal geodesic from e to k. Assume now that ζ is not horizontal. Then, by Lemma 5.1, we can find a horizontal curve $\zeta_h : [0,1] \to G$ with $\zeta_h(0) = k$ and $\zeta_h(1) \in G_b$ which is strictly shorter than ζ. Since ζ is a unique minimal geodesic between e and k, it cannot hold that $\zeta_h(1) = e$. But then we reach a contradiction, because it cannot then hold that $k \in K$, because there is a point $\zeta_h(1)$ on the identity fibre which is closer to k than e. Therefore, ζ must be horizontal, and so it descends to a corresponding geodesic $\bar{\zeta}$ between b and b'. $\bar{\zeta}$ must be unique minimal, otherwise ζ cannot be unique minimal. This finalises the proof. ☐

Remark 5.4. If the metric g_B is not right invariant, then Proposition 5.3 is still valid in the case $b'' = b$.

5.1 Optimal information transport

We now consider the main example in this paper, namely $G = \mathrm{Diff}^s(M)$ equipped with the α-β-γ metric (5) and $B = \mathrm{Dens}^{s-1}(M)$. As derived above, the α-β-γ metric descends to $\mathrm{Dens}^{s-1}(M)$, where it is given, up to multiplication with the constant β, by the Fisher metric. For simplicity, we assume throughout this section that $\beta = 1$. Recall that $\bar{R}_\varphi(\nu) = \varphi^*\nu$ and $\pi_{\mathrm{vol}}(\varphi) = \varphi^*\mathrm{vol}$. Also recall that if $s > n/2 + 1$, then $\mathrm{Diff}^s(M)$ and $\mathrm{Dens}^{s-1}(M)$ are Banach manifolds, and the projection $\pi_{\mathrm{vol}} : \mathrm{Diff}^s(M) \to \mathrm{Dens}^{s-1}(M)$ is smooth. Thus, all the prerequisites in Proposition 5.3 are fulfilled.

It was shown by Khesin, Lenells, Misiołek, and Preston [19] that the geodesic problem on $\mathrm{Dens}^{s-1}(M)$ with respect to the Fisher metric can be formulated as an optimal transport problem with respect to a degenerate cost function. However, as the cost function is degenerate, solutions are not unique, so there is no corresponding polarisation result. The α-β-γ metric on $\mathrm{Diff}^{s-1}(M)$ allows us to obtain a non-degenerate optimal transport formulation in accordance with the framework above. In particular, we obtain a factorisation result for diffeomorphisms.

Let $\lambda, \nu \in \mathrm{Dens}^{s-1}(M)$. The following Monge–Kantorovich problem is considered:

$$\text{Find } \varphi \in \{\phi \in \mathrm{Diff}^s(M); \phi^*\lambda = \nu\} \text{ minimising } \mathrm{dist}^2_{\alpha\beta\gamma}(\mathrm{id}, \varphi). \tag{14}$$

Here, $\mathrm{dist}_{\alpha\beta\gamma}$ is the Riemannian distance corresponding to the α-β-γ metric (5). Since the α-β-γ metric descends to the Fisher metric, we refer to (14) as *optimal information transport*.

Due to Proposition 5.3 it is enough to study geodesics on $\mathrm{Dens}^{s-1}(M)$ in order to solve (14). Also, since the α-β-γ metric is right invariant, it is no restriction to assume that $\lambda = \mathrm{vol}$.

As mentioned in the introduction, Friedrich [12] showed that the Fisher metric has constant curvature. This implies that its geodesics are easy to analyse. Indeed, following [19], we introduce the infinite dimensional sphere of radius $r = \sqrt{\mathrm{vol}(M)}$

$$S^{\infty,s}(M) = \Big\{ f \in \mathcal{F}^s(M); \langle f, f \rangle_{L^2} = \mathrm{vol}(M) \Big\}.$$

If $s > n/2$ then this set is an H^s Banach manifold. The L^2 inner product on $\mathcal{F}^s(M)$ restricted to $S^{\infty,s}(M)$ provides a weak Riemannian metric. Although weak, it has a

geodesic spray which is smooth. The geodesics are given by great circles, so it follows that $S_r^{\infty,s}(M)$ is geodesically complete and that its diameter is given by $\pi\sqrt{\mathrm{vol}(M)}$.

Let $\mathcal{O}^s(M) = \{f \in S^{\infty,s}(M); f > 0\}$ denote the space of positive functions of radius $\sqrt{\mathrm{vol}(M)}$. $\mathcal{O}^s(M)$ is an open subset of $S^{\infty,s}(M)$, and therefore a Banach manifold in itself. The following result is given in [19].

Theorem 5.5. *If $s > n/2$, then the map*

$$\Phi : \mathrm{Dens}^s(M) \ni \nu \longmapsto \sqrt{\frac{\mathrm{d}\nu}{\mathrm{dvol}}}$$

is an isometric diffeomorphism $\mathrm{Dens}^s(M) \to \mathcal{O}^s(M)$*. The diameter of* $\mathcal{O}^s(M)$*, and thus also* $\mathrm{Dens}^s(M)$*, is given by* $\frac{\pi}{2}\sqrt{\mathrm{vol}(M)}$*.*

Notice that if $f, g \in \mathcal{O}^s(M)$, then there is a unique minimal geodesic $\sigma : [0,1] \to S^{\infty,s}(M)$ from f to g, and that geodesic is contained in $\mathcal{O}^s(M)$, i.e., $\mathcal{O}^s(M)$ is a convex subset of $S^{\infty,s}(M)$. Indeed, the minimal geodesic is given by

$$\sigma : [0,1] \ni t \longmapsto \frac{\sin\big((1-t)\theta\big)}{\sin\theta} f + \frac{\sin(t\theta)}{\sin\theta} g, \quad \theta = \arccos\Big(\frac{\langle f, g\rangle_{L^2}}{\mathrm{vol}(M)}\Big). \tag{15}$$

The polar cone of $\mathrm{Diff}^s(M)$ with respect to the α-β-γ metric is given by

$$K^s(M) = \big\{\varphi \in \mathrm{Diff}^s(M); \mathrm{dist}_{\alpha\beta\gamma}(\mathrm{id}, \varphi) \le \mathrm{dist}_{\alpha\beta\gamma}(\phi, \varphi),\ \forall\, \phi \in \mathrm{Diff}^s_{\mathrm{vol}}(M)\big\}.$$

Since there exists a unique minimal geodesic between vol and any $\nu \in \mathrm{Dens}^{s-1}(M)$, it follows from Proposition 5.3 that every $\psi \in K^s(M)$ is the endpoint of a minimal horizontal geodesic $\zeta : [0,1] \to \mathrm{Diff}^s(M)$ with $\zeta(0) = \mathrm{id}$. Since ζ is horizontal, it is of the form $\zeta(t) = \mathrm{Exp}_{\mathrm{id}}(t\,\mathrm{grad}(w_0))$ for a unique $w_0 \in \mathcal{F}^{s+1}(M)$, where $\mathrm{Exp} : T\mathrm{Diff}^s(M) \to \mathrm{Diff}^s(M)$ denotes the Riemannian exponential corresponding to the α-β-γ metric.

Let $\varphi \in \mathrm{Diff}^s(M)$. Due to the explicit form (15) of minimal geodesics in $\mathcal{O}^{s-1}(M)$, and thus also $\mathrm{Dens}^{s-1}(M)$, we can compute the function $w_0 \in \mathcal{F}^{s+1}(M)$ such that $\mathrm{Exp}_{\mathrm{id}}(\mathrm{grad}(w_0))$ is the unique element in $K^s(M)$ belonging to the same fibre as φ. Indeed,

$$\begin{aligned} T\pi_{\mathrm{vol}} \cdot \mathrm{grad}(w_0) &= \frac{\mathrm{d}}{\mathrm{d}t}\Big|_{t=0} \pi_{\mathrm{vol}}(\mathrm{Exp}_{\mathrm{id}}(t\,\mathrm{grad}(w_0))) \\ &= \frac{\mathrm{d}}{\mathrm{d}t}\Big|_{t=0} \pi_{\mathrm{vol}}(\zeta(t)) = \frac{\mathrm{d}}{\mathrm{d}t}\Big|_{t=0} \sigma(t)^2 \mathrm{vol} = 2\dot\sigma(0)\sigma(0)\mathrm{vol}, \end{aligned}$$

where $\sigma(t)$ is the curve (15) with $f = 1$ and $g = \sqrt{\mathrm{Jac}(\varphi)}$ (recall that the Jacobian is defined by $\mathrm{Jac}(\psi)\mathrm{vol} = \psi^*\mathrm{vol}$). Since $\sigma(0) = 1$, and since $T\pi_{\mathrm{vol}} \cdot \mathrm{grad}(w_0) = \mathcal{L}_{\mathrm{grad}(w_0)}\mathrm{vol} = \Delta w_0 \mathrm{vol}$, we get

$$\Delta w_0 = \frac{2\theta\sqrt{\mathrm{Jac}(\varphi)} - 2\theta\cos(\theta)}{\sin\theta}, \quad \theta = \arccos\Big(\frac{\int_M \sqrt{\mathrm{Jac}(\varphi)}\mathrm{vol}}{\mathrm{vol}(M)}\Big). \tag{16}$$

Consider now the problem of finding the horizontal geodesic $\zeta(t) = \mathrm{Exp}_{\mathrm{id}}(t\,\mathrm{grad}(w_0))$. Of course, one can always solve equation (2) with $u(0) = \mathrm{grad}(w_0)$, and then reconstruct $\zeta(t)$ by integrating the non-autonomous equation $\dot\zeta(t) = u(t) \circ \zeta(t)$ with $\zeta(0) = \mathrm{id}$. However, since we know the projected geodesic curve (15) explicitly, we may also proceed by directly lifting that curve to a corresponding horizontal geodesic $\zeta(t)$. Indeed, the geodesic $\bar\zeta(t)$ in $\mathrm{Dens}^{s-1}(M)$ corresponding to $\zeta(t)$ is given by $\bar\zeta(t) = \Phi^{-1}(\sigma(t)) = \sigma(t)^2\mathrm{vol}$. Since $\pi_{\mathrm{vol}}(\zeta(t)) = \bar\zeta(t)$ and

$$TR_{\zeta(t)^{-1}}(\zeta(t), \dot\zeta(t)) = (\mathrm{id}, \dot\zeta(t) \circ \zeta(t)^{-1}) =: (\mathrm{id}, \mathrm{grad}(w_t)), \quad w_t \in \mathcal{F}^{s+1}(M)$$

we conclude that

$$\begin{aligned}\dot{\bar\zeta}(t) &= T_{\mathrm{id}}(\bar R_{\zeta(t)} \circ \pi_{\mathrm{vol}}) \cdot \mathrm{grad}(w_t) \\ &= \zeta(t)^*\big(\mathcal{L}_{\mathrm{grad}(w_t)}\mathrm{vol}\big) \\ &= \mathrm{div}(\mathrm{grad}(w_t)) \circ \zeta(t)\, \mathrm{Jac}(\zeta(t))\mathrm{vol}.\end{aligned}$$

By using $\dot{\bar\zeta}(t) = 2\dot\sigma(t)\sigma(t)\mathrm{vol}$ and $\mathrm{Jac}(\zeta(t)) = \sigma(t)^2$, we get

$$2\dot\sigma(t)\sigma(t) = \big(\Delta w_t \circ \zeta(t)\big)\sigma(t)^2.$$

The horizontal geodesic $\zeta(t)$, and its inverse $\zeta_{inv}(t) := \zeta(t)^{-1}$, can now be constructed by solving the following non-autonomous ordinary differential equation

$$\begin{aligned}\dot\zeta(t) &= \mathrm{grad}(w_t) \circ \zeta(t), \quad \zeta(0) = \mathrm{id} \\ \dot\zeta_{inv}(t) &= -\big(T\zeta(t)\big)^{-1} \cdot \mathrm{grad}(w_t), \quad \zeta_{inv}(0) = \mathrm{id} \\ \Delta w_t &= \frac{2\dot\sigma(t)}{\sigma(t)} \circ \zeta_{inv}(t) \\ \sigma(t) &= \frac{\sin\big((1-t)\theta\big)}{\sin\theta} + \frac{\sin(t\theta)}{\sin\theta}\sqrt{\mathrm{Jac}(\varphi)}, \quad \theta = \arccos\Big(\frac{\int_M \sqrt{\mathrm{Jac}(\varphi)}\mathrm{vol}}{\mathrm{vol}(M)}\Big).\end{aligned} \tag{17}$$

The equation for $\zeta_{inv}(t)$ is obtained by

$$0 = \frac{\mathrm{d}}{\mathrm{d}t}\Big(\zeta(t) \circ \zeta(t)^{-1}\Big) = \dot\zeta(t) \circ \zeta(t)^{-1} + T\zeta(t) \cdot \frac{\mathrm{d}}{\mathrm{d}t}\Big(\zeta(t)^{-1}\Big).$$

Notice that we already know that equation (17) has a unique solution for $t \in [0,1]$. As required, the equation is invariant with respect to any substitution $\varphi \to \phi \circ \varphi$ with $\phi \in \mathrm{Diff}^s_{\mathrm{vol}}(M)$.

In summary, we have proved the following result.

Theorem 5.6. *Let $s > n/2+1$. Then every $\varphi \in \mathrm{Diff}^s(M)$ admits a unique factorisation $\varphi = \phi \circ \psi$, with $\phi \in \mathrm{Diff}^s_{\mathrm{vol}}(M)$ and $\psi \in K^s(M)$. It holds that $\psi = \mathrm{Exp}_{\mathrm{id}}(\mathrm{grad}(w_0))$ with w_0 given by equation (16). There is a unique minimal horizontal geodesic $\zeta(t)$ with $\zeta(0) = \mathrm{id}$ and $\zeta(1) = \psi$, which can be computed by solving equation (17).*

Remark 5.7. Notice that the polar factorisation in Theorem 5.6 does not depend on the parameters α, β, and γ. The reason for this is that every parameter choice yields the same horizontal distribution, and the same horizontal geodesics.

5.2 Optimal transport of inner products and QR factorisation

In this section we show how the QR factorisation of square matrices is related to optimal transport of inner products on $\mathbb{R}^n$. The example provides a finite dimensional analogue to optimal information transport described in § 5.1 above. We do not rigorously address questions of global existence and uniqueness of geodesics (local existence and uniqueness follows automatically since we consider geodesics on a smooth manifold with respect to a smooth metric). Rather, the aim is to provide geometrical insight to the QR factorisation and its relation to the Cholesky factorisation.

The setting is as follows. Let $G = \mathrm{GL}(n)$ over the field $\mathbb{R}$ and let $B = \mathrm{Sym}(n)^+$ be the manifold of inner products on $\mathbb{R}^n$. $\mathrm{Sym}(n)^+$ is identified with the space of symmetric positive definite $n \times n$ matrices, i.e., if $M \in \mathrm{Sym}(n)^+$ is a symmetric positive definite matrix, then the corresponding inner product is $\langle \mathbf{x}, \mathbf{y} \rangle_M = \mathbf{x}^\top M \mathbf{y}$. Notice that $\mathrm{Sym}(n)^+$ is a convex open subset of the vector space $\mathrm{Sym}(n)$ of all symmetric $n \times n$ matrices.

The group $\mathrm{GL}(n)$ acts on $\mathrm{Sym}(n)^+$ from the right by $\bar{R}_A(M) = A^\top M A$. This action is transitive, which follows (for example) from the fact that every positive definite symmetric matrix has a Cholesky factorisation. If $U \in T_M\mathrm{Sym}(n)^+$, then the lifted action is $T_M \bar{R}_A \cdot U = A^\top U A$.

Let I denote the identity matrix (which is an element in both $\mathrm{GL}(n)$ and $\mathrm{Sym}(n)^+$), and consider the projection $\pi_I : \mathrm{GL}(n) \to \mathrm{Sym}(n)^+$ given by $\pi_I(A) = \bar{R}_A(I) = A^\top A$. The corresponding isotropy group is $G_I = \mathrm{SO}(n)$, which follows since $\pi_I(QA) = A^\top Q^\top Q A = A^\top A = \pi_I(A)$ for all $Q \in \mathrm{SO}(n)$. Thus, we have a principle $\mathrm{SO}(n)$–bundle $\pi_I : \mathrm{GL}(n) \to \mathrm{Sym}(n)^+$.

There is natural metric g_B on $\mathrm{Sym}(n)^+$ given by

$$\mathsf{g}_{B,M}(U, V) = \mathrm{tr}(U M^{-2} V), \quad U, V \in T_M\mathrm{Sym}(n)^+ = \mathrm{Sym}(n). \tag{18}$$

This metric is invariant with respect to the action $\bar{R}_A$. Indeed,

$$\begin{aligned}
\mathsf{g}_{B,\bar{R}_A(M)}(T_M\bar{R}_A \cdot U, T_M\bar{R}_A \cdot V) &= \mathrm{tr}(A^\top U A (A^\top M A)^{-2} A^\top V A) \\
&= \mathrm{tr}((A^\top M A)^{-1} A^\top U A (A^\top M A)^{-1} A^\top V A) \\
&= \mathrm{tr}(A^{-1} M^{-1} A^{-\top} A^\top U A A^{-1} M^{-1} A^{-\top} A^\top V A) \\
&= \mathrm{tr}(A^{-1} M^{-1} U M^{-1} V A) \\
&\text{(using cyclic property: } \mathrm{tr}(ABC) = \mathrm{tr}(BCA)) \\
&= \mathrm{tr}(M^{-1} U M^{-1} V A A^{-1}) \\
&= \mathrm{tr}(M^{-1} U M^{-1} V) \\
&= \mathrm{tr}(U M^{-2} V) = \mathsf{g}_{B,M}(U, V)
\end{aligned}$$

We now proceed by defining a metric on $\mathrm{GL}(n)$. Consider the projection operator $\ell : \mathrm{Mat}(n, n) \to \mathrm{Mat}(n, n)$ given by

$$\ell(U)_{ij} = \begin{cases} U_{ij} & \text{if } i > j, \\ 0 & \text{otherwise.} \end{cases}$$

Thus, $\ell(U)$ is the matrix which is equal to U at the strictly lower triangular entries, and zero at the other entries. Next, let g_G be the right invariant metric on $\mathrm{GL}(n)$ which at the identity is given by

$$\mathsf{g}_{G,I}(u,v) = \mathrm{tr}\left(\ell(u)^\top \ell(v)\right) + \mathrm{tr}\left((u+u^\top)(v+v^\top)\right). \tag{19}$$

By right translation we then have $\mathsf{g}_{G,A}(U,V) = \mathsf{g}_{G,I}(UA^{-1}, VA^{-1})$. Notice that the orthogonal complement of $T_I\mathrm{SO}(n) = \mathfrak{so}(n)$ with respect to this metric is given by the upper triangular matrices, which follows since matrices in $\mathfrak{so}(n)$ are skew symmetric, so the second term in (19) vanishes if either u or v belong to $\mathfrak{so}(n)$. In other words, $\mathfrak{so}(n)^\top = \mathfrak{upp}(n) := \{u \in \mathfrak{gl}(n); \ell(u) = 0\}$.

Proposition 5.8. *The right invariant metric g_G on $\mathrm{GL}(n)$ is descending with respect to π_I. The corresponding metric on $\mathrm{Sym}(n)^+$ is given by g_B.*

Proof. By Proposition 4.3 we first need to show that

$$\mathsf{g}_{G,I}(\mathrm{ad}_\xi(u), v) + \mathsf{g}_{G,I}(u, \mathrm{ad}_\xi(v)) = 0, \quad \forall\, u, v \in \mathfrak{upp}(n), \xi \in \mathfrak{so}(n).$$

We have

$$\mathsf{g}_{G,I}(\mathrm{ad}_\xi(u), v) = \mathrm{tr}\left(([\xi,u] + [\xi,u]^\top)(v+v^\top)\right) = \mathrm{tr}\left(([\xi, u+u^\top])(v+v^\top)\right).$$

By using the cyclic property of the trace we then get

$$\begin{aligned} \mathrm{tr}\left(([\xi,u+u^\top])(v+v^\top)\right) &= -\,\mathrm{tr}\left((u+u^\top)([\xi, v+v^\top])\right) \\ &= -\,\mathrm{tr}\left((u+u^\top)([\xi,v] + [\xi,v]^\top])\right) \\ &= -\mathsf{g}_{G,I}(u, \mathrm{ad}_\xi(v)). \end{aligned}$$

Therefore, the metric is descending. Next, if $u \in \mathfrak{upp}(n)$, then $T_I\pi_I \cdot u = u + u^\top$, so

$$\mathsf{g}_{G,I}(u,v) = \frac{1}{4}\,\mathrm{tr}\left((u+u^\top)(v+v^\top)\right) = \mathsf{g}_{B,I}(T_I\pi_I \cdot u, T_I\pi_I \cdot v).$$

Since g_B is right invariant it follows that g_G descends to g_B, which proves the result. ☐

The horizontal distribution $\mathcal{H}$ is given by $\mathcal{H}_A = \mathfrak{upp}(n)A$. Since $\mathfrak{upp}(n)$ is a Lie algebra, i.e., it is closed under the matrix commutator, it holds that the horizontal distribution is integrable. Its integral manifold through the identity is given by the Lie group of upper triangular $n \times n$ matrices whose diagonal entries are strictly positive. This Lie group is denoted $\mathrm{Upp}(n)$.

Let $A \in \mathrm{GL}(n)$. If there exists a unique minimal geodesic $\bar\zeta : [0,1] \to \mathrm{Sym}(n)^+$ from I to $\pi_I(A)$, then, in accordance with the framework above, we obtain a factorisation $A = QR$, with $Q \in \mathrm{SO}(n)$ and $R \in \mathrm{Upp}(n)$.

Remark 5.9. Since the metric (19) is smooth, it follows from standard results in Riemannian geometry that there exists a neighbourhood $\mathcal{O} \subset \mathrm{Sym}(n)^+$ of I such that any element in $\mathcal{O}$ is connected to I by a unique minimal geodesic. Therefore, if $\pi_I(A)$ is close

enough to I, then A has a unique QR factorisation. Also, the QR factorisation of any matrix is well known to exist, and is unique if the matrix is invertible, which suggests the existence of minimal geodesics. Details concerning these questions are not investigated in this paper.

In summary, we see that the factor R solves the problem of optimally (with respect to the cost function dist_G^2) transporting the Euclidean inner product on $\mathbb{R}^n$ to the inner product defined by $M = A^\top A$. Furthermore, the factor R is the transpose of the Cholesky factorisation of M. Indeed, if $L = R^\top$ then $M = \pi_I(A) = \pi_I(R) = R^\top R = LL^\top$.

Usually, the QR factorisation is obtain by direct linear algebraic manipulations. Another way to compute the R component is to solve the geodesic equation on $\mathrm{Sym}(n)^+$, and lift that geodesic to a horizontal geodesic on $\mathrm{GL}(n)$. Although this is probably inefficient compared to existing algorithms (there are very fast algorithms based on Householder reflections), the geodesic approach might nevertheless provide some insights, for example in the case of sparse matrices.

Remark 5.10. The setting can be extended to $\mathrm{GL}(n, \mathbb{C})$ by replacing $\mathrm{SO}(n)$ with $\mathrm{U}(n)$, and every transpose with the Hermitian conjugate.

Acknowledgements

The author would like to thank Martins Bruveris, Darryl Holm, Boris Khesin, Martin Kohlmann, and Matthew Perlmutter for very helpful discussions. The work is supported by the Royal Swedish Academy of Science, and the Swedish Research Council (contract VR–2012–335).

References

[1] S.-i. Amari and H. Nagaoka, *Methods of information geometry*, vol. 191 of *Translations of Mathematical Monographs*, American Mathematical Society, Providence, RI, translated from the 1993 Japanese original by Daishi Harada, 2000. 3

[2] V. I. Arnold, *Sur la géométrie différentielle des groupes de Lie de dimension infinie et ses applications à l'hydrodynamique des fluides parfaits*, Ann. Inst. Fourier (Grenoble) **16** (1966), 319–361. 2

[3] V. I. Arnold and B. A. Khesin, *Topological Methods in Hydrodynamics*, vol. 125 of *Applied Mathematical Sciences*, Springer-Verlag, New York, 1998. 2

[4] Y. Brenier, *Polar factorization and monotone rearrangement of vector-valued functions*, Comm. Pure Appl. Math. **44** (1991), 375–417. 3, 21

[5] N. N. Čencov, *Statistical decision rules and optimal inference*, vol. 53 of *Translations of Mathematical Monographs*, American Mathematical Society, Providence, R.I., translation from the Russian edited by Lev J. Leifman, 1982. 3

[6] G. E. Crooks, *Measuring thermodynamic length*, Phys. Rev. Lett. **99** (2007), 100602. 18

[7] D. G. Ebin and J. E. Marsden, *Groups of diffeomorphisms and the notion of an incompressible fluid.*, Ann. of Math. **92** (1970), 102–163. 9, 10, 11, 12, 17

[8] L. C. Evans, *Partial differential equations and Monge-Kantorovich mass transfer*, lecture notes, 2001. 19

[9] P. Facchi, R. Kulkarni, V. Man'ko, G. Marmo, E. Sudarshan, and F. Ventriglia, *Classical and quantum fisher information in the geometrical formulation of quantum mechanics*, Physics Letters A **374** (2010), 4801–4803. 18

[10] E. H. Feng and G. E. Crooks, *Far-from-equilibrium measurements of thermodynamic length*, Phys. Rev. E **79** (2009), 012104. 18

[11] R. A. Fisher, *Statistical methods and scientific inference*, Hafner Press [A Division of Macmillan Publishing Co., Inc.], New York, third edition, revised and enlarged, 1973. 3

[12] T. Friedrich, *Die fisher-information und symplektische strukturen*, Math. Nachr. **153** (1991), 273–296. 3, 22

[13] R. S. Hamilton, *The inverse function theorem of Nash and Moser*, Bull. Amer. Math. Soc. (N.S.) **7** (1982), 65–222. 17

[14] R. Hermann, *A sufficient condition that a mapping of Riemannian manifolds be a fibre bundle*, Proc. Amer. Math. Soc. **11** (1960), 236–242. 19

[15] D. D. Holm, T. Schmah, and C. Stoica, *Geometric mechanics and symmetry*, vol. 12 of *Oxford Texts in Applied and Engineering Mathematics*, Oxford University Press, Oxford, 2009. 2

[16] L. Kantorovich, *On a problem of monge*, Journal of Mathematical Sciences **133** (2006), 1383–1383, 10.1007/s10958-006-0050-9. 19

[17] L. Kantorovich, *On the translocation of masses*, Journal of Mathematical Sciences **133** (2006), 1381–1382, 10.1007/s10958-006-0049-2. 19

[18] B. Khesin, J. Lenells, and G. Misiołek, *Generalized Hunter-Saxton equation and the geometry of the group of circle diffeomorphisms*, Math. Ann. **342** (2008), 617–656. 3, 7

[19] B. Khesin, J. Lenells, G. Misiołek, and S. C. Preston, *Geometry of diffeomorphism groups, complete integrability and geometric statistics*, 2011. 3, 8, 16, 22, 23

[20] B. Khesin and R. Wendt, *The Geometry of Infinite-dimensional Groups*, vol. 51 of *A Series of Modern Surveys in Mathematics*, Springer-Verlag, Berlin, 2009. 2

[21] M. Kohlmann, *A note on multi-dimensional camassa–holm-type systems on the torus*, Journal of Physics A: Mathematical and Theoretical **45** (2012), 125205. 7

[22] S. Lang, *Fundamentals of differential geometry*, vol. 191 of *Graduate Texts in Mathematics*, Springer-Verlag, New York, 1999. 14

[23] J. Lenells, G. Misiołek, and F. Tiğlay, *Integrable evolution equations on spaces of tensor densities and their peakon solutions*, Comm. Math. Phys. **299** (2010), 129–161. 3

[24] J. E. Marsden and T. S. Ratiu, *Introduction to Mechanics and Symmetry*, vol. 17 of *Texts in Applied Mathematics*, second ed., Springer-Verlag, New York, 1999. 2

[25] R. J. McCann, *Polar factorization of maps on Riemannian manifolds*, Geom. Funct. Anal. **11** (2001), 589–608. 3

[26] R. J. McCann, *Five lectures on optimal transportation: geometry, regularity and applications*, lecture notes, 2010. 19

[27] G. Monge, *Mémoire sur la théorie des déblais et de remblais*, 1781. 19

[28] J. Moser, *On the volume elements on a manifold*, Trans. Amer. Math. Soc. **120** (1965), 286–294. 17

[29] F. Otto, *The geometry of dissipative evolution equations: the porous medium equation*, Comm. Partial Differential Equations **26** (2001), 101–174. 2, 19

[30] H. Poincaré, *Sur une forme nouvelle des équations de la mécanique*, C.R. Acad. Sci. **132** (1901), 369–371. 8

[31] C. Radhakrishna Rao, *Information and the accuracy attainable in the estimation of statistical parameters*, Bull. Calcutta Math. Soc. **37** (1945), 81–91. 3

[32] M. E. Taylor, *Partial Differential Equations. I*, vol. 115 of *Applied Mathematical Sciences*, Springer-Verlag, New York, basic theory, 1996. 3, 10, 13

[33] C. Villani, *Optimal transport*, vol. 338 of *Grundlehren der Mathematischen Wissenschaften*, Springer-Verlag, Berlin, 2009. 19

[34] C. Vizman, *Geodesic equations on diffeomorphism groups*, SIGMA Symmetry Integrability Geom. Methods Appl. **4** (2008), Paper 030, 22. 2